DOCUMENT RESUME

ED 363 501 SE 053 747

AUTHOR	Schmidt, Hans-Jurgen, Ed.
TITLE	Empirical Research in Chemistry and Physics Education. Proceedings of the International Seminar (Dortmund, Germany, June 10-12, 1992).
INSTITUTION	International Council of Associations for Science Education.
REPORT NO	ISBN-962-7532-06
PUB DATE	92
NCTE	197p.
PUB TYPE	Collected Works - Conference Proceedings (021)
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	Academic Achievement; *Chemistry; College Science; Concept Formation; Correlation; Educational Research; Elementary School Science; Elementary Secondary Education; Foreign Countries; Heat; Higher Education; High Schools; Meta Analysis; Misconceptions; *Physics; Piagetian Theory; *Problem Solving; Science Curriculum; *Science Education; Science Teachers; Scientific Concepts; Secondary School Science; Spatial Ability; Student Attitudes; *Student Evaluation; Textbooks; Thermodynamics
IDENTIFIERS	Analogies

ABSTRACT

In the course of the last decade the Dortmund Summer Symposium has developed into an internationally acknowledged conference, with English as the language of communication every other year. This book presents the following papers given at the 1992 conference: (1) Learning Science: Insights from Research on Teaching and Assessment (D. R. Baker); (2) Criterion-referenced assessment: A Case Study of the English National Curriculum (R. W. Fairbrother); (3) Strategies Students Use To Solve Chemistry Problems (M. L. Halpin); (4) The Educational Structure of Organic Synthesis (H. van Keulen); (5) Students' Abilities in Working with Chemistry Textbooks (G. Meyendorf); (6) Meta-Analytic and Multivariate Procedures for the Study of Attitude and Achievement in Science (M. D. Piburn); (7) Correlation and Causality (I. Pigeot); (8) "Work" and "Heat" in Teaching Thermodynamics (van Roon); (9) Methodological Aspects of University of Dortmund Studies of Students' Conceptions of Chemistry (E. Schach); (10) A Case Study on Students' Difficulties Applying the Bronsted Theory (H. Schmitt); (11) A Study of the Relationship between Conceptual Knowledge and Problem-Solving Proficiency (A. Shaibu); (12) Analogies in Senior High School Chemistry Textbooks: A Critical Analysis (R. B. Thiele, D. F. Treagust). Each paper is followed by a commentary of the discussant and a summary of the plenary discussion. (PR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ERIC

ED 363 501

ICASE

EMPIRICAL RESEARCH IN CHEMISTRY AND PHYSICS EDUCATION

Proceedings of the International Seminar, June 10th – 12th, 1992
University of Dortmund, Germany

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Dennis G. Chisman
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
[X] This document has been reproduced as received from the person or organization originating it
[ ] Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

SE053 747

Edited by HANS-JÜRGEN SCHMIDT

ERIC

BEST COPY AVAILABLE

Empirical Research in Chemistry and Physics Education

ICASE

ERIC

Hans–Jürgen Schmidt (Editor)

Proceedings of the International Symposium

EMPIRICAL RESEARCH IN CHEMISTRY AND PHYSICS EDUCATION

The International Council of Association for Science Education

ICASE

ERIC

Reproduction of papers in this seminar report does not preclude publication of the papers in other books or journals. Copyright permission to quote or reprint pages should be sought by the authors.

The papers and discussants' comments were printed as submitted. Authors are responsible for correctness of the individual texts and their contents.

ISBN-962-7532-06
Lofty Virtue Publication Centre

Editor: Prof. Dr. Hans-Jürgen Schmidt, University of Dortmund, Germany
Production: Dr. Heinz-Jürgen Kullmann, Wasserstraße 24, Dortmund, Germany

Reproduced in Hong Kong by the International Council of Associations for Science Education 1992.

Welcome

Once again I welcome all partcipants to this 3rd successful symposium on Empirical Research in Science and Mathematics Education. ICASE is very happy to endorse this symposium and to agree to print the proceedings. In this way ICASE hopes it can bring the interesting research information to the attention of teachers not only in Europe but worldwide.

This year the first time, discussions that took place during the symposium have also been included. I hope this makes the proceedings even more useful and enables ICASE to disseminate the outcomes even more widely than before. The proceedings have found their way into quite a number of libraries and I sincerely hope this trend will continue.

May I also take this opportunity to thank Professor Dr Hans-Jürgen Schmidt for all his efforts in making this symposium possible.

Dr Jack Holbrook
ICASE Excecutive Secretary

Content

PREFACE

Hans-Jürgen Schmidt, Universität Dortmund, Germany

In the course of the last decade the *Dortmund Summer Symposium* has developed into an internationally acknowledged conference, with English as language of communication every other year. The symposium is unique in that

- science education researchers and statisticians meet and co-operate at this conference and
- plenty of time is allowed for discussion between the individual talks.

Due to the amount of time spent on discussions only a limited number of speakers can be accepted. The conference will consequently never become a big event as to numbers of participants. But it has the great advantage that close contacts with other researchers in a kind of family atmosphere create new ideas and insights.

Two years ago *Mike Piburn* gave a very interesting paper as discussant at the Annual Meeting of the *National Association for Research in Science Teaching* after which it was suggested that he present his ideas at the Dortmund Symposium. When he sent the manuscript of his paper for the conference proceedings, the Dortmund statisticians *Iris Pigeot-Kübler* and *Elisabeth Schach* found Mike's paper perfectly in line with their own.

The reader will find that the individual researchers who came together for the Dortmund Symposium have used very different research methods. The summaries of the discussions, the contributions made by the discussants and the overall conference summary written by *Dale Baker* show to what extent ideas were interchanged and taken up.

Part of the texts were written by non-native speakers of English. Our Dortmund students who wrote the summaries of the discussions were kindly supported by English-speaking participants. I am very happy about their commitment.

When a Summer Symposium is over, I have a deep feeling of gratitude for all the effort, inspiration, enthusiasm and warmth that the participants brought forward. It is the human relations that make each conference so lively.

ERIC

Learning Science: Insights From Research on Teaching and Assessment

Dale R. Baker, Arizona State University, USA

1. Introduction

The purpose of this paper is to examine the major trends in science education in the areas of achievment, teaching and assessment and to place each of the papers presented at the 1992 Seminar on Empirical Research in Chemistry and Physics Education within the context of these trends. The trends to be examined are in the areas of (1) paradigms and methodologies, (2) measuring achievment, (3) understanding science (4) gender differences and (5) instruction.

2. Paradigms and Methodologies for Examining Teaching and Learning Paradigms

Cognitive psychology, especially constructivism, with special emphasis on prior knowledge, is the most influential paradigm in science education today. This paradigm has changed the concept of teaching and learning and the kinds of research questions we ask. For example, questions such as 'What kinds of concepts do students hold?' and 'What kinds of experiences form students' conceptions?' are the primary focus of much of the research in science education. These questions have, in turn, influenced the kind of research conducted. There is less process product research and experimental design comparing two groups (Gunstone, White, & Fensham, 1988) and more research grounded in the learners' conceptions of the world.

Closely related to the questions being asked about students are a new set of questions about teachers. These questions are also concerned with conceptions of the world as mediating factors in teachers' decision making processes. Questions such as 'How do teachers decide what to teach?', 'How do teachers' values effect the way they convey science?' and 'Where do teachers' explanations come from?' give insights into the way instructional decisions are made and effect how and what students learn. These questions have given rise to more qualitative ways of gathering and analyzing data and unlike the process product research place instructional strategies within the context of content, values and beliefs (Shulman, 1986).

The influence of Piaget has diminished and evolved into a Neo-Piagetian perspective. Neo-Piagetians have tried to blend Piagetian theory with information processing theory (Beilin, 1987) and in science education are particularly concerned with two factors not addressed by Piaget's theory. The first, is the mechanisms which facilitate the transition from one stage to the next and the second is individual differences.

ERIC

The majority of Neo-Piagetian research on transition mechanisms focuses on the theory of M-space proposed by Pascual-Leone. M-space is the amount of information an individual can process while solving problems (Niaz & Robinson, 1992; Pascual-Leone, Goodman, Ammon, & Subelman, 1978). The transition from one stage to the next is due to an increase of M-space and thus an increase in the amount of information that can be processed.

The research on individual differences focuses on disembedding as measured by field dependence/independence (Witkin, Moore, Goodenough, Cox, 1977) to explain differences in achievment (Lawson, 1983). Neo-Piagetians also question many of the Piagetian postulates, look for explanations for postulates such as the neuro-physiological evidence of self-regulation, build on ideas such as operative versus figurative knowledge (procedural vs. conceptual) and examine the influence of previous experience on science achievment[1].

3. Methodologies

The changes in research questions have led to more studies that use qualitative/ethnographic techniques for gathering data or a combination of both qualitative and quantitative techniques. Clinical and verbal interviews, classroom observations, discourse analysis, in-depth videotape analysis, examination of textbooks and curriculum material, journal writing & other reflective vehicles, case studies, concept maps and open ended and free response assessment instruments have become increasingly more popular. All of these approaches attempt to not only get at what is happening in the minds of teachers and students and in the classroom, but why. These techniques are labor intensive and time consuming but provide a depth of understanding not available from standardized assessments, but they can also limit the degree of generalizability of the research.

In direct opposition to the trend of in-depth analysis of small samples is the increase in large scale data gathering and secondary analyses such as meta-analysis. Increased computer power has made possible the International Assessment of Education Progress and nation-wide assessments. These assessments have allowed us to make valuable national and international comparisons. Meta-analysis has made it possible to aggregate data from research over many years and across multiple studies. These assessments provide a breadth of understanding and, in the case of meta-analysis, widespread generalizability.

These two trends provide a balanced body of research and increase our understanding of the teaching and learning of science on a micro and macro level.

Michael Piburn (1992), in his paper *Meta-analytic and multivariate procedures for identifying factors that are predictive of success in science*, provided evidence that large scale secondary analysis contributes to our knowledge in science education especially in regard to the Neo-Piagetian paradigm. According to his findings, the move away from this paradigm is warranted because M-space and field dependence/independence do not

[1]Conversation with A. Lawson, May 1992

appear to be measuring factors that are distinctly different from spatial ability and general ability (G). As such, M–space and field dependence/independence do not increase our ability to predict who will be successful in science.

Iris Pigeot–Kübler (1992), in her paper *Correlation and causality*, reminded us that with a more sophisticated approach and theoretical models to test, statistical analysis still has much to offer and can be a way to deal with the knowledge explosion in science education. She encouraged us to build and test theoretical models of possible relationships. This approach moves us away from a theoretical work to research that is based on previous scholarship within a theoretical framework. It also allows us to increase the generalizability of our findings and the predictive power of our research.

The research of Myra Halpin (1992), *Strategies students use to solve chemistry problems*; Hans–Jürgen Schmidt (1992), *A case study of students' difficulties applying the Brønsted Theory*; Peter van Roon (1992), *"Work" and "Heat" in teaching thermodynamics*; and Hanno van Keulen (1992), *Teaching organic synthesis in the laboratory* are all examples of the changes that have occurred in the questions we ask in our research in science education. They are also excellent examples of how the questions asked influence the methodology.

Myra Halpin (1992) used think aloud protocols for problem solving and interviews as well as paper and pencil assessments to examine patterns and correlates of different types of problem solvers. Hans–Jürgen Schmidt (1992) used error patterns in students' answers on standardized tests to guide his interviews which were then examined for patterns of misunderstandings.

Peter van Roon (1992) recorded students' discussions while they solved problems and he made extensive observations of students as they solved problems. These discussions and observations were then subjected to a content analysis by a group of scholars from a variety of backgrounds related to chemistry education.

Hanno van Keulen (1992), working within the paradigm called action research, studied his own teaching. He developed instructional strategies and materials that seemed most likely to work and then observed students as they worked in groups developing and testing procedures to create esters.

Elisabeth Schach (1992), in her paper *Methodological aspects of Dortmund studies of students' conceptions in chemistry*, emphasized the need for more careful research design in studies if we are to do the kinds of path–analysis discussed by Iris Pigeot–Kübler (1992) and the kind of secondary analysis presented by Michael Piburn (1992). I believe that some of the shift in methodology seen in the papers presented at this conference and education in general may be due to problems of design discussed by Elisabeth Schach (1992). Poor design leads to lack of confidence in the results of the study, which can, in turn, lead to the abandonment of an approach for another which is perceived to be more sound.

4. Measuring Achievment

4.1 Large Scale Assessment

International data on science achievment lead us to conclude that achievment is not as good as we would like it to be. Data from the 1989 International Assessment of Educational progress indicates that only students in Korea and the province of British Columbia in Canada can be called outstanding in mathematics and science. The range of performance was greatest for chemistry and the least for biology. Students were also assessed on their understanding of the nature of science. The data from all countries indicated that students did poorly, but Korean students did poorest of all (Baker, 1991).

These comparisons, when looked at on the surface, seem straight forward. Some countries are doing a better job than others at educating their students in the area of science, but no country is doing a really great job. However, these data have been called into question because the analysis did not control for differences in national resources, differences in time spent studying science in school, differences in the population of students taking the assessments and the difficulty in creating a test bank of questions that samples from a variety of national and international curricula.

Chemistry educators in particular, are not satisfied with standardized assessment as it presently stands. Students can do well on standardized tests using algorithms and formulae without understanding the concepts that underlie the problems they have solved. Students lack conceptual understanding and cannot qualitatively solve or represent problems. Consequently, chemistry educators are calling for a move toward assessing and teaching concepts (Ver Beck & Louters, 1991; Lythcott, 1990; Sawrey, 1990). The criticism and dissatisfaction with standardized testing have given rise to the movement for authentic assessment.

4.2 Authentic Assessment

At the same time that we have gone to large scale national and international assessment there has also been a move toward authentic assessment. Authentic assessment can be defined as evaluation that matches what is assessed more closely with (1) what is taught, (2) includes problem solving and process skills and (3) uses a wide range of techniques beyond standardized tests and multiple choice formats. Standardized tests, especially those using multiple choice formats, are not authentic because they cannot claim to closely match what is taught and are usually limited by the multiple choice format to knowledge, comprehension and application questions.

Authentic assessment questions large scale national and international assessment because the questions on such a test can not provide a good match with what is taught across all the participating states, provinces or nations. Even in situations in which there is a national or state curriculum local variation exists. In addition, these large scale efforts assess a limited range of fact and application knowledge and do not include questions on the more important problem solving and process skills.

Proponents of authentic assessment are suggesting that science educators replace standardized tests and multiple choice formats with a wide variety of assessment tools such as essays, practical assessments, portfolios, observations, interviews, concept maps, think aloud protocols and projects. They are also suggesting that assessment expands to include more summative evaluation and a greater emphasis on assessing the process of thinking and problem solving (Lawrenz, 1992).

All of the alternatives to standardized testing and multiple choice formats have the clear advantage of providing in-depth information about what a student has learned, how she/he solves problems and what misconceptions she/he holds. However, there are two disadvantages to authentic assessment. The first is that many of the approaches are time consuming and not every approach readily lends itself to assessing large numbers of students. The second is that the use of multiple assessors, say for the examination of laboratory skills, can lead to problems of inter rater reliability.

Many of the discussions concerning the research presented at this conference centered on what we mean when we say that we are measuring achievment in science. Some presenters have defined it as performance on a standardized test or a course grade, others as solving paper and pencil problems and another as performing real life laboratory activities. The variety of definitions at the conference is mirrored in the science education literature as illustrated by Michael Piburn's (1992) meta-analysis concerning factors that predict success in science.

Amos Shaibu, in his study *A study of the relationship between conceptual knowledge and problem-solving proficiency of science students in selected Nigerian schools: Pedagogical implications*, made it very clear that achievment is not a unified concept. Success in one area of science, conceptual knowledge, does not mean that students will be equally successful in another area of science, problem solving.

The work presented at this conference and the discussions that ensued lead one to conclude that achievment is a function of how it is measured and how it is measured often comes to define achievment. It is also true that how achievment is defined a priori can influence how it is measured.

Robert Fairbrother's (1992) paper, *Criterion-referenced assessment: A case study of the English National Curriculum*, is a good example of how legislative mandate based on a need for accountability and obvious political considerations runs counter to current thinking about assessment. He would not be faced with the problem of deciding on which questions are appropriate to measure a concept or deciding on the degree of difficulty the question should have if he were working in the authentic assessment paradigm. Even on a national scale, local administration allows for a variety of approaches that would provide a truer picture of student knowledge than that provided by standardized testing. On the other hand, a national assessment such as that being implemented in the United Kingdom which imposes constraints that limits testing to three one hour sessions, makes authentic assessment difficult if not impossible.

Many of the techniques in the authentic assessment repertoire such as observations, interviews, and practical assessments are also found in the realm of research. These techniques have been used successfully by David Treagust (1992) who interviewed textbook authors in his study *Analogies in senior high school chemistry textbooks*, Hans-Jürgen Schmidt (1992) and Myra Halpin (1992) who interviewed students, Peter van Roon (1992) who observed students and Hanno van Keulen (1992) who used practical assessments.

This represents, in my opinion, a convergence of concerns in the school and research community. Both have acknowledged that the true measure of learning is understanding and that understanding is not measured accurately by the kinds of standardized tests that presently exist.

5. Understanding Science

5.1 The Nature of Science

Concern about students' understanding of the nature of science is another trend in science education arising, in part, from recent national and international assessments that indicate that students worldwide have a poor understanding of the nature of science (Baker, 1991). Although understanding the nature of science has been a goal of science education for many years, little time or space in the curriculum has been allocated to the nature of science (Bybee, Ellis, Mathews, 1992). Since assessment in the United States using the National Assessment of Educational Progress (National Assessment of Educational Progress, 1989) and international tests now include items on the nature of science, instruction will have to be broadened.

Instruction about the nature of science is important for its own sake, but even more important is the role an understanding of the nature of science has in bringing about conceptual change. Duschl (1991) provides a convincing argument that how a student understands the nature of science can be critical to whether the student relinquishes a misconception and embraces a scientifically more accurate concept.

However, teaching about the nature of science is more difficult than teaching facts or even concepts. First, teachers lack the background in the history, philosophy and sociology of science which is needed to teach about the nature of science. Second, textbooks allocate little space to the history and development of scientific ideas, the struggles over ideas in the history of science or the application of science to students' lives. Although the research on the influence of teachers' conceptions of science on students' conceptions of science is inconclusive, there is evidence that teachers' conceptions do influence how and what they teach. Teachers who understand the nature of science encourage more higher level thinking and frequently use problem solving, inquiry oriented instruction and higher level questioning in a risk free environment (Ledeman, 1992).

Understanding the nature of science was not central to the research presented at this conference. Yet, there is strong evidence that it is an important factor in relinquishing misconceptions and an understanding of the nature of science can also influence how teachers teach.

I would encourage researchers to broaden the way they look at students to include how they understand the nature of science as a way to shed light on the source of student misconceptions and problem solving strategies. Rules of evidence and ways of knowing that do not reflect the nature of science are as likely to cause problems as errors in factual knowledge and heuristics.

ERIC

5.2 Reasoning and Problem Solving

Most students have the potential to solve problems in science, but very few do it well. We know that culture does not affect reasoning and problem solving but context does. These contextual factors include such things as task familiarity and the number of hours spent studying science (Baker, 1991).

We also know that problem solving is not a dichotomy in which you either can or can not solve problems, but a continuum from novice to expert. Experts understand that problems require careful analysis and reasoning. They use better heuristics, have more flexibility, greater knowledge and can qualitatively represent problems. Novices, on the other hand, see problems as tasks which can be solved in one or two steps. They use poorly remembered formulae, lack basic knowledge such as translating chemical symbols into words and have a superficial level of problem representation. Novices are careless and they don't check their work (Baker, 1991).

Amos Shaibu's (1992) research indicated that although the students in his study had an excellent knowledge of facts and concepts, this knowledge was not sufficient for successful problem solving. The research in problem solving emphasizes the need for extensive knowledge, but it is clearly a necessary but not sufficient condition for success.

Myra Halpin (1992) examined the relationship of choice of problem solving strategies as a function of spatial and personality variables. Further research in this area might include a look at qualitative representations of problems and other characteristics of experts which are amenable to change through instruction rather than the more fixed characteristics of personality.

Both researchers might consider rethinking how they assess problem solving and abandon the simple dichotomy of correct or incorrect solutions. Instead they should examine the degree of correctness as a better representation of a continuum of skills.

Hanno van Keulen's (1992) work is a good example of how important context specific instruction is for problem solving. He provided a strong argument for the need to move away from instruction in a general approach to solving problems, e. g. recognizing similarities and differences or interpreting data, to instruction in problem solving that is based on a specific context and draws upon specific prior knowledge and skills.

5.3 Misconceptions

This is an area in which there has been a great deal of research in the last ten years. This research has given rise to new ways of diagnosing students' understanding of scientific concepts. For example, researchers are now using case studies, interviews and tests in which students give both the answer and their reasons for the answer with the reasoning being the more important of the two. These techniques have their antecedents in the clinical interviews of Piaget. Misconception researchers are also indebted to Piaget for the influence of his theory on their guiding questions; how conceptions are formed and how they can be molded (Gunstone, White, & Fensham, 1988).

Researchers have cataloged the variety of student held misconceptions and concluded that misconceptions are extensive, pervasive and quite similar from country to country. There is also a large body of evidence to suggest that traditional science instruction does

ERIC

little to change misconceptions and sometimes increases them (Osborne & Wittrock 1983).

Out of these two bodies of research has come the idea of the conceptual change models of instruction. The models emphasize the use of instructional strategies designed to help students restructure their knowledge so that their theories and the meaning of concepts they hold more closely resemble those of the scientist.

There are several models each of which proposes ways in which instruction can be structured to help students relinquish misconceptions and embrace more scientificly accurate concepts. All of the models, at minimum, consist of two propositions; (1) the learner must experience dissatisfied with his/her existing conception and (2) the learner must find the new conception intelligible, plausible and fruitful (Posner, Strike, Hewson, & Gertzog, 1982). However, the translation of the various models into instruction has not brought about large changes in how students view the world. This is due less to flaws in the models and more to the resistance to change of deep seated and long held ideas (Gunstone, White, & Fensham 1988).

Duschl and Gitomer (1991) present a conceptual change model that emphasizes ongoing assessment that provides many opportunities to confront and challenge students' knowledge and includes as knowledge students' epistomological frameworks. In the model, students are taught to assess the quality of their conceptual and epistomological ideas according to established scientific criteria.

Duschl and Gitomer argue that changes must first occur in a student's epistomological framework in order for conceptual change to take place. The student's epistomological framework influences what she/he considers evidence for a scientific explanation and consequently whether the explanation will be embraced. Many researchers (Strike & Posner, 1992; Duschl & Gitomer, 1991; Carey, Evans, Honda, Jay & Unger, 1989) believe that the failure to asses and confront epistomological frameworks at variance with those used in science may account for some of the lack of success researchers have had in bringing about conceptual change.

Duschl and Gitomer also contend that conceptual change models necessitate changes in the role of the learner and teacher, and changes in the view of science and the goals of the curriculum. Addressing misconceptions within the old framework of teachers' and students' roles and the traditional curriculum will not lead to conceptual change. Since the discussions which followed the papers of Peter van Roon (1992) and Hans-Jürgen Schmidt (1992) included questions concerning what teachers can do to help students abandon misconceptions, I would suggest a careful review of the research in this area. Many scholars have examined the classroom conditions, student characteristics, topics and time needed for the successful implementation of a conceptual change model of instruction.

Peter van Roon's (1992) work is concerned with students' understanding of science and raises the question of the source of the lack of understanding. His work may be viewed as part of the research in misconceptions, but unlike many who work in this area, he does not take a psychological perspective. Instead of examining pre-existing schemes, he looks at semantic or language problems as the source of misconceptions. This approach raises two questions for me. Is changing the language used by teachers and students sufficient for changing conceptual understanding? and Isn't it possible for students to manipulate the new vocabulary with the same facility as they manipulated the old and still not understand? Conceptual change models imply that much more than a change in vocabulary must take place before there is a change in how students view the world.

Hans–Jürgen Schmidt's (1992) research is more reflective of the work done in the area of misconceptions in that he identified, through interviews, the reasons why students chose distractors rather than correct answers on a multiple choice test. He identified two factors, the application of a wrong theory and confusion about electron and proton transfer that can help teachers when they teach about the Brønsted Theory.

6. Gender Differences

6.1 Achievment

Gender differences in science achievment favoring males were found in the 1989 International Assessment of Educational Progress for Korea, Spain, Canada and Ireland but not for the other participating countries. Research on gender differences in achievment world wide sometimes favor boys and sometimes girls (Linn & Hyde, 1989; Stromquist, 1989). For example, gender differences favoring males are not present in Thailand where females have been found to do better than males in chemistry, physics and math. The preponderance of evidence, such as the small size and instability of gender differences, indicates that the differences are not general but specific to the context or culture in which the studies have taken place. They are reflection of cultural values and the expectations for girls to study science rather than innate differences between males and females (Baker, 1991).

6.2 Rates of Participation

What is more important than looking at gender differences in achievment is looking at gender differences in rates of participation. When we talk about girls' achievment in science we are still talking about a very small portion of all students in school from primary through tertiary level. In the third world, where the situation is the worst, women's enrolment even in primary and secondary school lags behind men's in all areas except Latin America and the Caribbean.

In most third world countries, women are kept out of school to help with the domestic chores of the family and to engage in income earning activities. Parents, especially in Moslem countries, do not see the education of girls as worth the cost and even view education as interfering with their daughters' marriageability. Only in cases where families are from high socio–economic levels and the parents are educated will girls attend university. Even then, women tend to cluster in traditional female areas of study such as nursing and teaching.

Schools in third world countries do little to reverse these trends. Textbooks and curriculum materials reinforce traditional sex roles. Teachers have been found to have little regard for females' ability and counsellors do not provide females with information about nontraditional careers. Governments, families and schools are least supportive of women's education in predominantly Muslim countries (Stromquist, 1989).

An examination of gender differences was not included in any of the conference papers except for the one presented by Myra Halpin. However, many scholars believe that gender should be part of any research design where feasible. Gender should be included, not to simply to find differences between males and females, which tells us nothing about underlying causes, but to determine whether analogies, textbooks and teacher behaviors work against female participation and success in science. Since males and females bring different experiences to the classrooms because of the way they have been socialized, one can not assume that what works for males will automatically work for females and vice versus.

Even if some of the participants of this seminar believe that gender differences are not a problem in his/her country, in most developing countries women do not receive a scientific education. Thus, as science educators and participants at an international seminar gender issues should be a concern for us all.

7. Instruction

7.1 Learning From Text

Reading about science is no substitute for doing science yet for many reasons both good and bad textbooks will always be with us. That being the case, the challenge facing science educators is to identify what constitutes good text and how and when to use textbooks to the maximum advantage. To begin with, it is important to understand the process of reading comprehension. Reading researchers do not view readers as passive recipients of information that resides in the text. Reading comprehension is an interactive process in which meaning is constructed by readers as a function of their prior knowledge. When the students' knowledge, such as a misconception, contradicts the information found in the text, the students' prior knowledge can and often does prevail over the information found in the text (Dole, Duffy, Roehler & Pearson, 1991). For example, Boyle & Maloney (1991) found that most students failed to use textual information about Newton's third law in solving physics problems even when they remembered reading that the text stated explicitly that all forces are equal.

In addition, comprehension strategies such as drawing inferences and meta–cognitive skills such as comprehension monitoring will also influence what students learn from textbooks (Dole, Duffy, Roehler & Pearson, 1991). Otero and Camapanario (1990) using chemistry and physics text, found that few (18.9 %) of tenth grade students and a little more than half (58 %) of twelfth grade students had the meta–cognitive skills to accurately evaluate their own text comprehension. Thus, the effectiveness of a science text is dependent upon the same set of factors that affect other instructional strategies used in science.

Gerhard Meyendorf's research, *Students' abilities in using chemistry school books*, examined how students gain information from text under three conditions. In condition one students were told what passages to read, in condition two students were told to find the appropriate passages and information for themselves and in condition three students were instructed in how to use the text to find information. This research in embedded in an older reading paradigm, the skills approach, which although it has merits, has been superseded by a more cognitively oriented paradigm with better explanatory power. Many of the results of Gerhard Meyendorf's study such as poor note taking or not reading the summaries can be explained by the newer constructivist perspective. Whether a student reads the summaries can be explained as a function of their meta-cognitive skills on comprehension monitoring and the quality of notes taken as the effect of existing schema on the assimilation of new information.

Further work with science textbooks should look to reading research for guidance, especially the work in schema theory, in regard to misconceptions. Other factors that influence a student's ability to extract information from text such as the placement of questions and the cognitive demands of integrating textual and graphical or pictorial information should also be examined.

7.2 Analogies

One line of research that science education scholars have begun to pursue is an examination of how texts are written. In particular, whether the inclusion of analogies in text will increase comprehension and counter the effect of prior conceptions. The research in this area is contradictory, in part because this is a new area of investigation, but also because metaphors, similes and examples are included in the definition of analogies

Gilbert (1989) found that using analogies as a general literary device did not effect achievment or recall and had a negative effect on attitude. Rather than helping clarify science concepts by providing a bridge for understanding the analogies further confused the students.

Brown (1992) found that a textbook excerpt that contained multiple examples analogous to the assessment question (Newton's third law) were less effective than a written text which contained a series of connected analogies which started from an anchoring example and lead to the target problem. Interviews with students indicated that the multiple textbook examples were counter intuitive while those which were explicitly connected were not. The student could not see the connection of the textbook examples with the target problem but could see the connection with the analogies which were written to emphasize connections. He concluded that examples that teachers find compelling may not be compelling to students. Despite the teachers perceptions, students may not see the examples as analogous to the target problems.

David Treagust's (1992) work is more precise in the definition of analogies than some of the research in this area. Precision in defining analogies and distinguishing analogies from metaphors and similes will go a long way to clarify our understanding of the role of analogies in teaching science.

David Treagust's (1992) work revealed that the use of analogies in textbooks is not under the control of science educators in terms of whether they will be included or whether the analogies are effective or well written. The authors of the science textbooks were outside of the science education community and were unaware of any model for teaching with analogy. I wonder if the collaborative writing of textbooks in which scholars in science education, content experts, and reading experts would not result in better textbooks and therefore more learning? I also believe that the objections raised by textbook authors concerning the inclusion of analogies are spurious. The inclusion of a few analogies per chapter can not possibly result in a loss of flexibility for teachers, but merely provide a stating place for instruction.

7.3 Pedagogical Content Knowledge

Closely related to teachers' ability to implement conceptual change instructional strategies is their pedagogical content knowledge. The idea of pedagogical content knowledge emerged in the late 1980's from Shulman's Knowledge Growth in Teaching Project at Stanford (Shulman, 1986) which grew out of a need to look beyond teachers' knowledge of subject matter and or teachers' knowledge of pedagogy.

Pedagogical content knowledge can be defined as how teachers relate their knowledge of content and their knowledge of pedagogy to the needs of specific learners in specific situations. It is not the blending of content and pedagogy but an understanding of what makes learning of specific content difficult and what instructional strategies or ways of representing the knowledge (analogies, demonstrations, explanations, models) make the knowledge understandable. Consequently, the teachers' science knowledge acquired during university instruction must be reorganized to represent the perspective of teaching rather than the research perspective of a scientist.

Pedagogical content knowledge also includes knowledge of (1) what motivates students, (2) students' attitudes toward science, learning and school, (3) the cognitive development and reasoning abilities of a wide range of students, (4) students' conceptions of themselves and science and (5) the preconceptions of students of different ages and backgrounds which they bring to learning science. In addition, it includes knowledge of the cultural, social, political and physical environments in which students learn. The teacher's choice of pedagogy takes all of these factors into consideration.

Pedagogical content knowledge is not so much the quality and content of a teacher's knowledge of his/her subject, pedagogy, students and context but how this knowledge is used effectively. Pedagogical content knowledge is characterized by and requires the restructuring of knowledge to fit the needs of students and the ongoing changes in the depth of a teacher's understanding of the teaching and learning process. Consequently, it is not readily accessible to novice teachers and is more often found in expert teachers (Cochran 1992).

The implications for teacher preparation and science teaching are great because content knowledge must go beyond facts and concepts. It must include understanding the structure of the subject matter; the way concepts and principles are organized and the way truth or falsehood is established. It includes both knowledge of scientific laws and the reasons why the laws are true. It includes knowledge of why certain ideas are central to a discipline, such as the periodic table in chemistry, and why others are not.

Pedagogical knowledge must also include a broad range of instructional techniques, knowledge of many different kinds of curricular materials and how each of the techniques and materials work with particular learning difficulties.

It is a teacher's pedagogical content knowledge that enables him or her to effectively bring about conceptual change. Diagnosing misconceptions and choosing instructional strategies and experiences that will compel a student to relinquish a long held conception and embrace another more scientifically accurate conception requires extensive knowledge of content, pedagogy, student characteristics and contextual factors.

Teachers' misunderstanding of analogies and their failure to use them, as reported by David Treagust (1992) indicates that the teachers in his study have poor pedagogical content knowledge. If teachers cannot identify the best analogies based on knowledge gained through both formal preparation to teach and experience gained in the classroom they cannot really be considered good teachers.

However, teacher preparation programs cannot be expected to provide teachers with a list of the best analogies to always use because the effectiveness of an analogy is context specific. What works for advanced students may not work for beginners and what works for native speakers may not be as effective for students with language problems. On the other hand, teacher preparation programs can help teachers restructure their own knowledge and teach them how to assess student characteristics so that they can identify for themselves the best analogies to use in a specific context.

Hanno van Keulen's (1992) work fits well within the framework of research in pedagogical content knowledge. In fact, I believe that his method of instruction and empirical approach to deciding how to teach can both improve the pedagogical content knowledge of the university instructor and provide a learning environment, through the modelling of good practice, that can improve the pedagogical content knowledge of prospective high school teachers.

8. Conclusion

In conclusion, I would like to say that research in the teaching and learning of science benefits most when researchers from a variety of perspectives come together to present data, discuss issues and exchange views. This seminar has provided us with such an opportunity. As a consequence, we have a clearer understanding of different methodologies, theoretical orientations, questions and concerns. This understanding has stimulated discussion and helped us, as researchers, to clarify and improve our own work.

References

Baker, D. (1991). *A summary of research in science education – 1989*. Science Education, vol. 75 (3), 255 – 402.)

Beilin, H. (1987). *Current trends in cognitive development research: Towards a new synthesis*. In B. Inhelder, D. de Caprona & A. Cornu-Wells (Eds.), Piaget Today (pp. 37 – 64). United Kingdom: Lawrence Erlbaum Associates Ltd.

Boyle, R. K. & Maloney, D. P. (1991). *Effects of written text on usage of Newton's third law*. Journal of Research in Science Teaching, vol. 28, 123 – 139.

Brown, D. E. (1992). *Using examples and analogies to remediate misconceptions in physics: Factors influencing conceptual change*. Journal of Research in Science Teaching, vol. 29 (1), 17 – 34.

Bybee, R. W., Ellis, J. D. & Mathews, M. R. (1992). *Teaching about the nature of science and* *technology: An introduction*. Journal of Research in Science Teaching, vol. 29 (4), 327 – 329.

Carey, S., Evans, R., Honda, M., Jay, E. & Unger, C. (1989). *"An experiment is when you try it and see if it works": A study of grade 7 students' understanding of the construction of scientific knowledge*. International Journal of Science Education, voi. 11, 1123 – 1130.

Cochran, K. F. (1992). *Pedagogical content knowledge: Teachers' transformations of subject matter*. In: F.Lawrenz, K. Cochran, J. Krajcik & P. Simpson (Eds.). Research Matters to the Science Teacher, 5 (pp. 3 – 10). Manhattan KS: NARST.

Dole, J. A., Duffy, G. G., Roehler, L. R. & Pearson, P. D. (1991). *Moving from the old to the new: Research on reading comprehension and instruction*. Review of Educational Research, vol. 61 (2), 239 – 264.

Duschl, R. A. & Gitomer, R. H. (1991). *Epistomological perspectives on conceptual change: Implications for educational practice*. Journal of Research in Science Teaching, vol. 28 (9), 839 – 858.

Fairbrother, R. (1992). *Criterion-referenced assessment: A case study of the English National Curriculum*. Paper presented at the Seminar on Empirical Research in Chemistry and Physics Education, Dortmund, Germany.

Gilbert, S. W. (1989). *An evaluation of the use of analogy, simile, and metaphor in science texts*. Journal of Research in Science Teaching, vol. 26 (4), 315 – 327.

Gunstone, R., White, R., & Fensham P. (1988). *Development in style and purpose of research on learning of science*. Journal of Research in Science Teaching, vol. 25 (7), 513 – 529.

Halpin, M. (1992). *Strategies students use to solve chemistry problems*. Paper presented at the Seminar o Empirical Research in Chemistry and Physics Education, Dortmund, Germany.

Lawrenz, F. (1992). *Authentic assessment*. In: F. Lawrenz, K. Cochran, J.Krajcik & P. Simpson (Eds.), Research Matters to the Science Teacher, 5 (pp. 65 – 70). Manhattan KS: NARST.

Lawson, A. (1983). *Predicting science achievment: The role of developmental level, disembedding ability, mental capacity, prior knowledge and beliefs*. Journal of Research in Science Teaching, vol. 22 (2), 117 – 129.

Ledeman, N. (1992). *Students' and teachers' conception of the nature of science: A review of the research*. Journal of Research in Science Teaching, vol. 29 (4), 331 – 359.

Linn, M & Hyde, J. (1988). *Gender, mathematics and science*. Educational Researcher, vol. 18 (8), 17 – 19, 22 – 27.

Lythcott, J. (1990). *Problem solving and requisite knowledge in chemistry*. Journal of Chemical Education, vol. 67, 248 – 252.

Meyendorf, G. (1992). *Students' abilities in using chemistry school books*. Paper presented at the Seminar on Empirical Research in Chemistry and Physics Education, Dortmund, Germany.

National Assessment of Educational Progress, (1989). *Science Objectives*. Princeton, NJ: Educational Testing Service.

Niaz, M. & Robinson, W. R. (1992). *Manipulation of logical structures of chemistry problems and its effect no student performance*. Journal of Research in Science Teaching, vol. 29 (3), 211 – 226.

Osborne, R. J. & Wittrock, M. C. (1983). *Learning Science: A generative process*. Science Education, vol. 67 (4), 489 – 508.

Otero, J. C. & Camapanario, J. A. (1990). *Comprehension evaluation and regulation in learning science from texts*. Journal of Research in Science Teaching, vol. 27 (5), 447 – 460.

Pascual-Leone, J., Goodman, D., Ammon, P., & Subelman, I. (1978). *Piagetian theory and neo-piagetian analysis as psychological guides in education*. In: J. M. Gallagher & J. A. Easley (Eds.), Knowledge and Development, New York: Plenum.

Piburn, M. (1992) *Meta-analytic and multivariate procedures for identifying factors that are predictive of success in science.* Paper presented at the Seminar on Empirical Research in Chemistry and Physics Education, Dortmund, Germany.

Pigeot-Kübler, I. (1992). *Correlation and causality.* Paper presented at the Seminar on Empirical Research in Chemistry and Physics Education, Dortmund, Germany.

Posner, G. J., Strike, K. A., Hewson. P. A., & Gertzog, W. A. (1982). *Accommodation of a scientific conception. Toward a theory of conceptual change.* Science Education. vol. 66, 211 - 227.

Sawrey, B. A. (1990). *Concept learning versus problem solving: Revisited.* Journal of Chemical Education, vol. 67, 253 - 254.

Schmidt, H.-J. (1992). *A case study of students' difficulties applying the Brønsted theory.* Paper presented at the Seminar on Empirical Research in Chemistry and Physics Education, Dortmund, Germany.

Schach, E. (1992). *Methodological aspects of Dortmund studies of students' conceptions in chemistry.* Paper presented at the Seminar on Empirical Research in Chemistry and Physics Education, Dortmund, Germany.

Shaibu, A. (1992). *A study of the relationship between conceptual knowledge and problem-solving proficiency of science students in selected Nigerian schools: Pedagogic implications.* Paper presented at the Seminar on Empirical Research in Chemistry and Physics Education, Dortmund, Germany.

Shulman, L. S. (1986). *Those who understand: Knowledge growth in teaching.* Educational Researcher, vol. 15 (2), 4 - 14.

Strike, K., & Posner, G. (1992). *A revisionist theory of conceptual change.* In: R. Duschl & R. Hamilton (Eds.), Philosophy of Science, Cognitive Psychology, and Educational Theory and Practice. Albany: SUNY Press.

Stromquist, N. P. (1989). *Determinants of educational participation and achievement of women in the third world: A review of the evidence and a theoretical critique.* Review of Educational Research, 59, (2), 143 - 183.

Treagust, D. (1992). *Analogies in senior high school chemistry textbooks: A critical analysis.* Paper presented at the Seminar on Empirical Research in Chemistry and Physics Education, Dortmund, Germany.

Van Keulen, H. (1992). *Teaching organic synthesis in the laboratory.* Paper presented at the Seminar on Empirical Research in Chemistry and Physics Education, Dortmund, Germany.

Van Roon, P. (1992). *"Work" and "heat" in teaching thermodynamics.* Paper presented at the Seminar on Empirical Research in Chemistry and Physics Education, Dortmund, Germany.

Ver Beck, K. & Louters, L. (1991). *Chemical language skills, investigating the deficit.* Journal of Chemical Education, vol. 68, 389 - 392.

Witkin, H. A., Moore, C. A., Goodenough, D. R. & Cox, P. W. (1977). *Field-dependent and field-independent cognitive styles and their educational implications.* Review of Educational Research, vol. 47 (1), 1 - 64.

CRITERION–REFERENCED ASSESSMENT: A CASE STUDY OF THE ENGLISH NATIONAL CURRICULUM

Robert W. Fairbrother, University of London, United Kingdom

Abstract

This paper deals with the problems in a criterion–referenced national assessment system of obtaining valid and reliable measures of attainment. It deals mainly with the reliability of standard assessment tasks which are composed for all Year 9 pupils (age 14) and concentrates in attainment in knowledge and understanding which is one of the two main components of the English national curriculum in Science; there is no discussion of attainment in investigations which is the other major component in science. The paper is in three parts;

- the first summarises the national curriculum in England and Wales so that there is an understanding of the circumstances in which the research took place;
- the second gives a brief discussion of reliability and of the main factors influencing the results of a trial during the development of standard assessment tasks;
- the third looks in detail at some of the results of the trial and discusses the consequences.

1. The National Curriculum in England and Wales

Provision for the establishment of a *National Curriculum* in England and Wales is set out in the Education Reform Act, 1988 which became law in schools in September 1989. The detailed provisions of the National Curriculum are given in statutory instruments supported by non–statutory guidance and various circulars. The main source of information for teachers in schools is provided by subject documents which give much of the detail which has to be taught in each subject. This paper makes particular use of the subject documents in Science (DES 1989, 1991).

An additional and very important determinant of change is the work of the *Task Group on Assessment and Testing* (DES 1988). This group set out the general framework for assessment in the *National Curriculum*. Central to its recommendations was the belief that assessment should be an integral part of the educational process, and that it should be the servant, not the master, of the curriculum.

A major aspect of the new curriculum is to have a coherent scheme which enables pupils to progress to their maximum level of achievement from age 5 when compulsory education starts to age 16 when it finishes. Reports of progress of pupils have to be made to

ERIC

parents each year. Aggregated results of schools have to be reported more widely at certain key stages. Progress is reported in ten levels of attainment, each of which is defined by statements of attainment. These definitions represent a move to a criterion-referenced system of assessment.

This paper reports some of the results of an attempt to measure pupil performance in science using the system described briefly above. In order to understand the work which was done it is necessary to know something about the new system which is being introduced into schools in England and Wales, and so the paper starts with a description of the national curriculum.

1.1 Subjects and profile components

There are ten compulsory Foundation subjects plus religious education. Three of the foundation subjects have special importance and are called *Core Subjects*. See Table 1.

Table 1: Compulsory subjects for all pupils aged 5 to 16

Core subjects
English, Mathematics, Science
Other Foundation subjects
Art, Foreign Language (from age 11), Geography, Technology (including design), History, Music, Physical Education

It is expected that virtually all pupils in key stage 4 will take the core subjects for the GCSE.

In an attempt to describe a subject more meaningfully, and to give more information about pupil performance, each subject is divided into Profile Components (PCs), and the Profile Components are divided into *Attainment Targets* (ATs). At the time of writing Science, for example, has two PCs *Exploration of Science*, which consists of one AT, and *Knowledge and Understanding of Science*, which consists of 16 ATs. Changes to this structure are proposed, (DES, 1991), but the general principles remain the same and do not affect the main issues reported in this research. The attainment of each pupil will be reported in each Profile Component as well as in the subject as a whole. Both the old system and the proposed new system for Science are summarised in Table 2 and ex-

Table 2: The structure of science in the National Curriculum

Old System					**New System**				
PROFILE COMPONENTS					PROFILE COMPONENTS				
1.	2.				1.	2.			
Exploration of science	Knowledge and understanding				Scientific investigation	Knowledge and understanding			
AT1	AT2	AT3	. . .	AT17	AT1	AT2	AT3	AT4	AT5
— Statements of attainment (SoAs) —					— Statements of attainment (SoAs) —				

plained further below.

1.2 Attainment targets

Profile Components are divided into *Attainment Targets* (ATs) which are the knowledge, skills and understanding which pupils are expected to have. Table 3 shows the proposed new attainment targets for science.

Table 3: The proposed new attainment targets in science

1	Scientific investigation
2	Life and living processes
3	Earth and environment
4	Materials and their behaviour
5	Energy and its effects

As an indication of what is meant by these attainment targets, a description of them is given in Table 4.

Table 4: A description of the new attainment targets

SCIENCE

Attainment Target 1 — Scientific investigation

Pupils should develop the intellectual and practical skills that allow them to explore the world of science and to develop a fuller understanding of scientific phenomena, the nature of the theories explaining these, and the procedures of scientific investigation. This work should take place in the context of activities that require a progressively more systematic and quantified approach, which draws upon an increasing knowledge and understanding of science. The activities should encourage the ability to plan and carry out investigations in which they:

(i) hypothesise and predict
(ii) observe and measure
(iii) interpret results and draw inferences
(iv) evaluate scientific evidence

2. Life and living processes

Pupils should develop their knowledge and understanding of:

i) the organisation of living things and of the processes which characterise their survival
ii) the diversity and classification of life-forms including the causes of variation and the basic mechanisms of inheritance, selection and evolution
iii) the factors affecting population size and human influences within ecosystems
iv) energy flows and cycles of matter within ecosystems.

3. Earth and environment

Pupils should develop their knowledge and understanding of:

i) the Earth, its weather and atmosphere
ii) the structure and resources of the earth
iii) the range of energy sources and the principles of thermal efficiency
iv) the Earth's place in the universe

ERIC

Table 4 (continuation): A description of the new attainment targets

4. Materials and their behaviour Pupils should develop their knowledge and understanding of: i) the properties, classification and structure of materials ii) the processes by which materials are changed by chemical reactions to form new materials iii) the behaviour of materials
5. Energy and its effects Pupils should develop their knowledge and understanding of the nature of energy through its transmission and transfer. They should develop their understanding through a study of: i) forces ii) electricity and electromagnetic effects iii) wave motion as illustrated by the properties and behaviour of light and sound.

The five proposed new attainment targets combine together most of what was in the old attainment targets. The descriptions for all the attainment targets are quite broad. This breadth is necessary since the targets apply to all pupils and so have to cover a wide range of ages and ability. They must be broad enough to allow pupils to attain different levels within each target. A narrow, precisely defined target would not permit a range of performances. More detailed information about the attainment targets is given by *Statements of Attainment*. These are explained later.

1.3 Key stages and progression

The 11 years of compulsory education are divided into four key stages as shown in Table 5.

Table 5 The key stages of compulsory education

Key stage	Ages	Phase	
1	5 – 7	Primary	
2	7 – 11	Primary	Middle
3	11 – 14	Secondary	Middle
4	14 – 16	Secondary	

1.3.1 Levels of attainment

Pupil progression through the subject from age 5 to age 16 is measured in ten levels of attainment. Each level provides a signpost for the next; a step which represents the average educational progress of children over about two years. Reports to parents have to be made each year but information about the level attained only has to be given at the end of each key stage. There will, of course, be a spread of attainment with some pupils at different levels from others. Typically pupils should be capable of achieving around the levels shown in Table 6 at or near the reporting ages of 7, 11, 14 and 16.

Table 6: Average level of attainment in a subject at the end of each key stage

Age	Average level
7	2
11	4
14	5 – 6
16	6 – 7

Successive levels in an attainment target are defined by statements of attainment, and the main purpose of assessment is to help pupils to make progress. The statements of attainment for levels 1, 3 and 5 in the proposed new *Attainment Target 4* are shown in Table 7. The even–numbered levels have been omitted in order to save space, however, looking at levels 1, 3 and 5 enables the intended progression to be seen more easily.

The statements of attainment are the main factors in deciding the level which a pupil has attained. Many people think of them as simple criteria which, on their own, enable reliable judgements to be made and common standards to be achieved. They can certainly help to do these things but decisions also have to take into account the age of the pupils being considered. For example, when interpreting the meaning of "explain the physical differences between solids, liquids and gases in simple particle terms" in level 5, it is necessary to make judgements as to what to expect of typical 13–year old pupils. The criteria cannot be interpreted in isolation from the norms of what it is reasonable to expect of pupils. The need to use judgements in this way means that the assessment of pupils is not 100 % reliable. This is no different from assessments which have been made in the past, say for the GCSE, and will be made in the future. All measurements are unreliable to some degree.

Table 7: The statements of attainment in levels 1,3 and 5 of Attainment Target 4

	Attainment Target 4 – Materials and their behaviour Pupils should:
Level 1 A typical 5–year old pupil	a) be able to identify familiar and unfamiliar objects in terms of their simple properties
Level 3 A typical 9–year old pupil	a) know that some materials occur naturally while many are made from raw materials
Level 5 A typical 13–year old pupil	a) be able to classify aqueous solutions as acidic, alkaline or neutral using pH. b) understand how to separate and purify the components of mixtures using physical processes. c) understand simple oxidation processes, including combustion, as reactions with oxygen to form oxides. d) be able to explain the physical differences between solids, liquids and gases in simple particle terms.

ERIC

1.4 Assessing and reporting

The level, from 1 to 10, achieved in an attainment target is determined by success in the statements of attainment which define that level. Information about performance in statements of attainment comes from teacher assessments which take place continually as a part of normal teaching, and from *Standard Assessment Tasks* (SATs) which are devised centrally but administered and marked by teachers in the school "at or near the end of the key stage". The SATs are used in schools in the first half of the Summer term at the end of the key stage.

The results of the SATs have to be combined with those from the teacher assessments. For the first statutory assessments at end of key stage 1 it was required (DES, 1990a) that the teacher assessments be reported in about April, and that any discrepancies between teacher assessments and *Standard Assessment Tasks* be resolved by an appeal process. This procedure was changed for the following year and enabled the teacher assessments and the SAT results to be reported at the same time, when the SATs had been marked by the teachers. A report has to be made available about each pupil by July 31st at the end of key stages 1, 2 and 3, and by 30th September at the end of key stage 4. (DES, 1990b)

Rules are applied to enable the *Attainment Target Levels* to be determined from performance in the statements of attainment and then aggregated to give *Profile Component Levels*, which in turn are aggregated to give subject levels. These requirements are summarised in Table 8 for key stages 1 – 3.

Table 8: Summary of assessment requirements

End of key stages 1 – 3						
Teacher assessment (Continuous)	+	Standard Assessment Tasks (Between April and June)	=	Final Report (End July)		
Arriving at levels						
Statements of Attainment	–	Attainment Target levels 1 – 10	–	Profile Component levels 1 – 10	–	Subject levels 1 – 10

The details are being changed as more experience is being gained about the functioning of the national curriculum, but the changes do not affect the general principles. The detailed arrangements for key stage 4 are not yet known.

2. Reliability

2.1 Introduction

This brief section clarifies the meaning of reliability and identifies two main factors, the organisation and administration of the SATs and the interaction of the pupils with the activities, which influence reliability in relation to the standard assessment tasks. There is also a summary of the methods which were used to measure reliability.

2.2 Meaning of reliability

2.2.1 Absolute reliability

Reliability is a continuous measure and so it is not possible to say that one measure is reliable and another is unreliable. All that can be said is that one measure may be more reliable than another. A decision as to whether a measure is sufficiently reliable is difficult to make, and relies a great deal on the professional judgement of the composers and users of a test. The decision is complicated by two major aspects of reliability which are set out below.

2.2.2 Reliability as repeatability of scores or levels

This is concerned with the extent to which we can rely upon the level decided for a pupil in an attainment target, profile component or subject being repeated if the pupil were assessed on another occasion using the same SATs or with a different but parallel set of SATs. Two major factors which influence reliability are the organisation of the SATs to ensure consistent procedures in different schools, and the interaction between the tasks and the pupils which gives rise to consistent performances in SoAs, ATs and PCs.

A common system of organisation in different schools helps to reduce variability of performance caused by such things as different teacher/pupil interaction and different assessment standards. In the work done to develop the SATs the major ways in which some commonalty was achieved were:

(a) guidance to teachers in the administration of the SATs

(b) assessment guides to help teachers towards a common interpretation of pupil responses.

The details of these are not given here.

The interaction between the tasks and the pupils can be explored by seeing how pupils perform in different measures. Consistent performances in SoAs, ATs and PCs can be checked by assessing the same thing several times. This is the main concern of this paper and is covered in section 3 below.

2.2.3 Reliability as homogeneity

Pupils may perform differently in the assessment of the same SoA or AT in different tasks depending on the different attributes being assessed. Many SoAs, and most levels within an AT, are defined by more than one attribute. (See Table 7 above). This means most SoAs and ATs lack homogeneity, and pupils may perform differently depending upon which attribute of the SoA or level in the AT is being assessed. Performance may also be affected by the context in which the assessments are placed. A different context arises when pupils are presented with a different activity usually on a different occasion.

If different facility values are obtained for different measures, this could point to unreliability of the measures or to a different grasp by pupils of different attributes of the same SoA or AT.

2.3 Measuring reliability

Conventional measures of reliability use correlation coefficients between two or more assessment occasions. The criterion–referenced information obtained from the SATs does not easily lend itself to correlation analyses which were developed to deal with data which is normally distributed. Furthermore, we were particularly interested in the extent to which different measures give the same levels of attainment; correlation coefficients only show that there is a relationship between the levels, not that the levels are the same. We therefore adopted the following procedures:

- SoA scores were treated as dichotomous data (yes/no or 1/0) and were analysed using facility values (F–values). We looked at the F–values of measures under two circumstances:
  - (a) when a statement of attainment or attainment target is assessed more than once in the same context;
  - (b) when a statement of attainment or attainment target is assessed in different contexts.
- AT levels (and PC and subject levels) obtained from aggregated SoA scores were considered to be putting pupils in ranks which can be dealt with using rank order correlation analyses.

It is not sufficient, however, to consider just the numerical results of statistics. If similar F–values are produced, one must use professional judgement to decide whether this is because the range of attributes or contexts covered was too narrow thus reducing the validity. If different F–values are produced, professional judgement must be used yet again to decide whether the same SoA or AT is being assessed but involving different attributes or different contexts.

3. Some detailed results

The chain of data gathering for the national curriculum assessment and reporting is:

$$\text{SoA} \rightarrow \text{AT} \rightarrow \text{PC} \rightarrow \text{Subject}$$

We were concerned about producing valid and reliable measures of attainment in attainment targets using standard assessment tasks sent to schools. However, performance in attainment targets depends crucially on performance in the statements of attainment which make up the attainment targets. An analysis of the measures of performance in SoAs was thus very important.

3.1 The data

The trial involved approximately 10,000 pupils in 100 schools in the summer term 1990. Different activities were composed each of which aimed to assess pupil performance in levels 3 to 7 in a particular attainment target. Some of the activities involved only written work and some involved a mixture of written work and practical work. The activities were given to the pupils by their teachers during the normal science time–table in the schools. We asked the teachers not to treat the activities as examinations but to try to

ERIC

use them as if they were a part of their normal teaching. The pupils' responses were written in booklets which we supplied, and the teachers marked the responses using assessment guides which we also supplied. The teachers received some instruction in the purpose and organisation of the exercise, and in interpreting the questions and assessment guides. Appendix A shows one of the activities, *Holiday Centre*, together with its assessment guide. This activity is discussed in detail later in this section.

Schools were able to choose activities according to certain rules so as to make up a balanced package of SATs. In order to try out the effect of different combinations of activities we had several different packages which we called *Models*. Altogether there were five models but pupils in any one school tried only one model.

This section looks at the results of assessing statements of attainment in levels 3 to 7 in the old *Attainment Target 13*, energy, in *Models 1B* and *1C*. The statements of attainment in all the levels of AT13 are shown in Appendix B. Table 9 shows the relevant activities in the two models. The activities were identified by names which gave some indication of the context of the activity.

Table 9: The activities in Models 1B and 1C

Model 1B	Model 1C
Holiday Centre	Energy Sources
Batlins	Heating
Boiler	Energy Story
	Home Comfort
	Heat Transfer

The data associated with *Models 1B* and *1C* were used to check on the reliability of some statements of attainment and attainment targets.

The main problems addressed were:

- how reliable are the measures of the statements of attainment?
- how stable are measures across different contexts, i. e. different activities?
- how reliable are the measures of the attainment targets?

3.2 Reliability of SoA measures

Some statements of attainment were assessed more than once in the same activity. Sometimes these assessments were similar to each other in what they asked the pupils to do and so can be used to give some indication of the repeatability of the assessment of these statements of attainment. Sometimes the assessments were different from each other and so can be used to give information more about the homogeneity of the attainment target at that level than about the repeatability of its assessment. Some statements of attainment were assessed in different activities, and so can give information about the stability of the assessments in different contexts.

As mentioned in section 2.3 above, the basic method of analysis is to compare F–values of different measures. If the different measures are assessing the same thing, one would

ERIC

expect the same or similar F–values. It would be reasonable to expect F–values for measures of different SoAs in the same level to be closer to each other than to those of SoAs in different levels. In addition, one would expect to see a pattern where the F–values at the lower levels are higher than those at the higher levels. However, it is necessary to emphasise what has been said earlier about needing to use professional judgement to interpret the statistics

3.2.1 Model 1B

In *Model 1B* the SoAs which were assessed in AT13 are shown in Table 10. (The letter 'a' refers to the first statement in the level, 'b' to the second, and so on. See appendix B.)

Table 10: Statements of Attainment assessed in AT13 in Model 1B (The numbers in parentheses shows the number of times the SoA was assessed)

Level	Holiday Centre	Batlins	Boiler
3	a(x4)		b(x1)
4	a(x1), b(x1), c(x2)	b(x1), c(x1)	d(x1)
5		a(x3), b(x1)	a(x4), b(x3)
6	a(x1), b(x1)		a(x3), b(x2)
7		a(x10), c(x1)	b(x2)

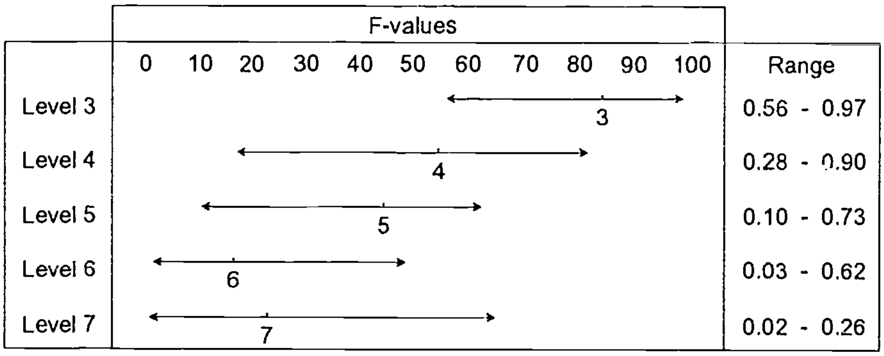
The teachers used the assessment guides to help them decide whether or not the responses from the pupils indicated performance at a particular level. They recorded their decisions as ticks or crosses in small boxes on the question paper. (See appendix A). The question papers were then returned to the SAT developers who transferred the decisions into a computer as 1 or 0. This information was obtained for a sample of between 960 and 1020 pupils who took *Model 1B*, for all the statements of attainment shown in Table 10 above. (The fluctuation in the number of pupils is caused by such things as absences from one day to the next and omission of parts of the tasks.) Table 11 on the next page shows the F–values for measures of individual SoAs, for the sum of all the measures of the same SoA (e. g. All a), for all measures at the same level in the same task (Total), and for all measures at the same level (All).

Table 11: F–values for the SoAs assessed in AT13 in Model 1B (Decimal points omitted)

Level	Holiday Centre		Batlins		Boiler		All
3	a_1 78, a_2 56	All a 82					
	a_3 97, a_4 97				b 93	All b 93	84
		Total 82				Total 93	
4	a 33	All a 33					
	b 28	All b 28	b 71	All b 71			
	c_1 64, c_2 54	All c 59	c 45	All c 45	d 90	All d 90	55
		Total 45		Total 58		Total 90	
5			a_1 73, a_2 55, a_3 51	All a 60	a_1 51, a_2 60, a_3 71, a_4 40	All a 56	
			b 37	All b 37	b_1 10, b_2 15, b_3 34	All b 20	45
				Total 54		Total 41	
6	a 49	All a 49			a_1 09, a_2 07, a_3 23	All a 13	
	b 03	All b 03			b_1 07, b_2 20	All b 14	17
		Total 26				Total 13	
7			a_1 21, a_2 18, a_3 05, a_4 07	All a 15			
			a_5 39, a_6 07, a_7 35, a_8 09		b_1 02, b_2 41	All b 22	
			a_9 03, a_{10} 04				20
			c_1 62, c_2 33	All c 48			
				Total 18		Total 22	

It can be seen that the F–values for the measures within any one level vary quite widely. This information is summarised in Figure 1 on the next page.

Figure 1: Spread of F–values in each of the levels of AT13 in Model 1B



3.2.2 Model 1C

Similar information was obtained for AT13 in *Model 1C*, and so we can see whether the same variability exists wit. another group of pupils but this time in a different context. The SoAs which were assessed are shown in Table 12.

Table 12: SoAs assessed in AT13 in Model 1C (The numbers in parentheses shows the number of times the SoA was assessed.)

Level	Energy Sources	Heating	Energy Story	Home Comfort	Heat Transfer
3	a		a		
4		b, d, e (x2)	b, c		
5				a, b	
6			a, b, d		
7				a, b, c	a (x2)

As in the case of *Model 1 B* the F–values obtained for a sample of about 1000 pupils who took *Model 1C*, for all the SoAs in Table 12, were obtained. These are shown in Table 13.

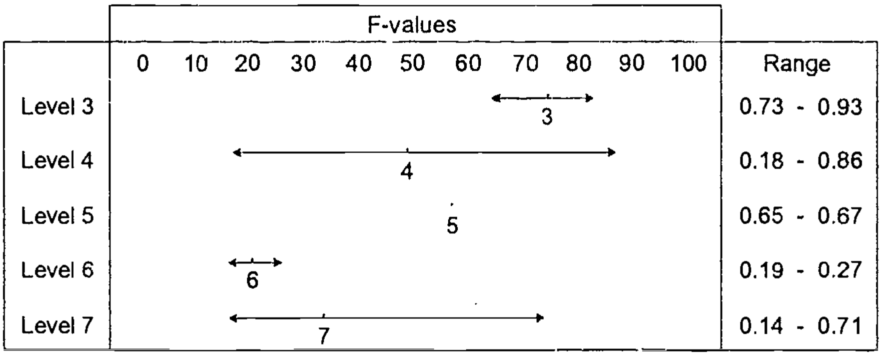
There is a tendency for the F–values within a level to be rather more homogeneous than in *Model 1B*, but the variation is still quite large. The range of F–values in each level is illustrated in Figure 2.

In the most extreme case (level 4 in *Model 1C*) choosing the hardest measure as an indicator of attainment rather than the easiest would affect nearly 70 % of the pupils. The smallest range is 2 % in level 5 of *Model 1C*, but typical values show a range of at least 20 %. Before drawing any conclusions from these analyses it is necessary to look in more detail at the questions and the assessment guides which were used to obtain the data.

Table 13: F-values for the SoAs assessed in AT13 in Model 1C (Decimal points omitted)

Level	Energy Sources		Heating		Energy Story		Home Comfort		Heat Transfer		All
3	a 93	All a 93 Total 93			a 73	All a 73 Total 73					85
4			b 67 d 86 e_1 e_2 52 49	All b 67 All d 86 All e 51 Total 64	b 18 c 27	All b 18 All c 27 Total 23					52
5							a 67 b 65	All a 67 All b 65 Total 66			66
6					a 27 b 19 d 24	All a 27 All b 19 All d 24 Total 23					23
7							a 71 b 14 c 40	All a 71 All b 14 All c 40 Total 42	a_1 a_2 22 19	All a 21 Total 21	33

Figure 2: Spread of F–values in each of the levels of AT13 in Model 1B



3.2.3 The questions and the assessment guides

We must be satisfied that the questions and assessment procedures give reliable and valid measure of the SoAs. For this purpose the assessments of 13/3a (AT13, level 3, SoA a) and of 13/4c (AT13, level 4, SoA c) which were done in the Holiday Centre activity in *Model 1B* are discussed in some detail below. Table 14 shows the relevant F–values extracted from Table 11.

Table 14: The F–values for the SoAs in levels 3 and 4 of the Holiday Centre activity in Model 1B

Level	Holiday Centre		
3	a_1 78	a_2 56	All a 82
	a_3 97	a_4 97	
4	a 33		All a 33
	b 28		All b 28
	c_1 64	c_2 54	All c 59

The assessments of 13/3a

It can be seen in Table 14 that SoA 13/3a was assessed four times in the *Holiday Centre* activity. Statement of attainment 13/3a says (see appendix B):

- understands, in qualitative terms, that models and machines need a source of energy in order to work.

This is quite a broad construct, and achievement will depend upon what is meant by 'understand', and upon which particular models and machines are being considered. There is a large number of possible assessments within which could be a range of performances all of which might be indicative of performance at this level.

The similarity of the results for assessments 3 and 4 indicate that they are each covering the same aspect which may be different from that covered by assessments 1 and 2. The first two assessments required the pupils to choose from a set of pictures four examples of where energy is being used and where the energy comes from. (See appendix A) We thought this was clear and straightforward and, together with the illustrative example given in the question, capable of being understood by pupils at the end of key stage 3 who are operating at level 3.

The pupils were required to respond in writing. For 14 year–old pupils operating at level 3 this may not be easy, and the assessment guide was written to try to take account of this difficulty. We expected pupils to be able to give one example and match it correctly with an energy source. In addition we thought that if pupils could give either another match of example with energy source or give examples and sources without necessarily matching them correctly, this would also give valid evidence for performance at level 3. Although this second alternative seems less rigid in its requirements the facility value is, in fact, lower. This could be for a variety of reasons, for example some of the teachers may have omitted to give credit for the second alternative if the pupils were given it for the first.

It is a matter of professional judgement whether these are valid assessments of this SoA. After reviewing the responses of the pupils and relating them to the assessment guidelines, we believe activities 1 and 2 are valid assessments of 13/3a and that the two F–values would be closer if stronger assessment guidelines were given.

The second two assessments of 13/3a required the pupils to explain what would happen to the sailing boat (a) if the wind blows harder, and (b) if the wind stops blowing. The SoA asks for an understanding, and to satisfy this the question asks for an explanation. The assessment guidelines suggest acceptance of statements of fact (the boat goes faster; the boat slows down or stops) without requiring an explanation. This, almost certainly, accounts for the high facility value of these two assessments since giving an explanation of the facts would be more difficult. It is almost certain that an assessment guideline which gave credit for an explanation would reduce the facility values of these two assessments and might give a greater agreement with the performances in assessments 1 and 2. Giving an explanation would be a more valid assessment of the understanding required for 13/3a, and one would be looking for a performance which it would be reasonable to expect from an average 9 year–old pupil who had been taught about this area of science.

The assessments of 13/4c

Both of the 13/4c assessments ask the pupils to fill in the gaps in sentences which refer to energy transfers taking place. The first is concerned with a personal stereo, and the second with swimming across a pool. As Table 14 shows, the first assessment is slightly easier with 64 % of the pupils getting it right compared with 54 % in the second assessment.

Statement of attainment 13/4c says:

- ❑ understand that energy can be stored, and transferred to and from moving things.

As with 13/3a, this is a relatively broad construct and the discussion which took place for 13/3a is relevant here. After considering the responses of the pupils our judgement is that these two activities and their assessment guidelines are valid assessments of 13/4c, but cover different attributes.

In the assessment guidelines we tried out different amounts of guidance. Generally speaking, however, we wanted to see to what extent teachers would be able to interpret the SoAs and their pupils' responses without a lot of detailed guidance. The variation in F–values results partly from this policy since the teachers were only just becoming familiar with the national curriculum, with the interpretation of the statements of attainment and with matching pupil performance to the statements.

This detailed look at the assessments of 13/3a and 13/4c has revealed opportunities for increasing the reliability of measures of SoAs, mainly through giving teachers more guidance in interpreting pupil responses. Similar consideration given to other assessment occasions in the trial reinforces this conclusion.

3.3 Reliability between different contexts

Models 1B and *1C* both assessed AT13 in different contexts. That is to say, the pupils were presented with different activities which were signalled by issuing different material often in different lesson periods. (Limitations of space mean that these activities cannot be shown here.) We can see in Tables 11 and 13 that there is variability in performance both within and between the different contexts. The variability between contexts at the same level can be seen by looking at the mean F–value in each context at each level (the total values in Tables 11 and 13). This information is shown separately in Table 15.

Table 15: Total F–values for each context in AT13 in Models 1B and 1C (Decimal points omitted)

Level	Model 1B			Model 1C				
	Holiday Centre	Batlins	Boiler	Energy Sources	Heating	Energy Story	Home Comfort	Heat Transfer
3	82		93	93	73			
4	45	58	90		64	23		
5		54	41				66	
6	26		13			23		
7		18	22				42	21

These values are the weighted means of the separate measures and make use of all the information which is available at each level. There is not sufficient data to be able to do an analysis of variance which would give some indication of whether the variation within the contexts is significantly different from that between the contexts. However, simple inspection of the data shows that the range of F–values *within* the contexts is generally greater than that *between* contexts (roughly a range of 40 within compared with 20 between). This indicates that the first priority is to achieve reliable measures within one context. However, the variability across contexts also points to the need to have more than one context.

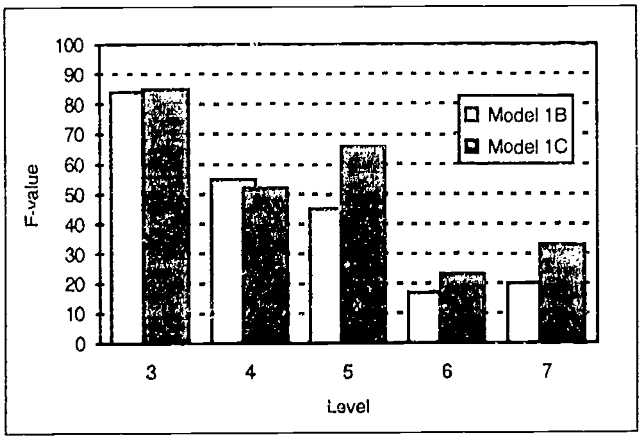
3.4 Reliability of attainment target measures

Some indication of performance in attainment targets can be obtained by combining all the measures within a level to form a mean F–value for the level. This is shown in the *All*– column of Tables 11 and 13 for the two assessments of AT13. These values are reproduced in Table 16, and plotted in Figure 3.

Table 16: Mean F–values for all the measures of AT13

Level	Model 1B	Model 1C
3	84	85
4	55	52
5	45	66
6	17	23
7	20	33

Figure 3: Mean F–values for all the measures in each level of AT13 in Model 1B and Model 1C



We do not know whether these are the right patterns (or nearly the right patterns). They will be influenced by a variety of factors including the representativeness of the schools involved. Schools were allocated to models so as to give a similar distribution of schools by socio–economic grouping in each model, but we cannot be certain that this distribution is representative of the population of schools in the country.

Keeping this limitation in mind we would expect to see a more evenly stepped pattern in both of the models. The pattern for *Model 1B* shows that discrimination between levels is achieved in the way one would expect with the exception of level 6 which seems to be rather low. In *Model 1C* different pupils were involved in different tasks and inconsistencies start earlier.

We can see a possible reason for these discrepancies when we look at the nature of AT13 at these levels and the nature of the tasks which were used. (See appendix A) The statements of attainment for levels 5, 6 and 7 in AT13 contain one strand which is concerned with energy economy, efficiency and conservation. In addition level 6 contains an

ERIC

SoA dealing with machines, and level 7 contains an SoA dealing with conduction, convection and radiation.

Dealing first with *Model 1B*, the tasks which gave information about level 7 concentrated on the SoA dealing with conduction, convection and radiation, while the tasks for level 6 concentrated on energy economy, efficiency and conservation. At this stage in the teaching of the national curriculum both pupils and teachers would find it easier to cope with and interpret questions about energy transfer than about energy economy etc. We believe that this is the main reason for the discrepancy and that it will gradually disappear as greater familiarity is gained with the national curriculum and in particular with the interpretation of the SoAs at different levels.

This interpretation is supported in *Model 1C* where a similar discrepancy occurs between levels 6 and 7 with tasks which dealt with the same SoAs as in *Model 1B*. The particularly high performance in level 5 comes from only two measures in one context both of which arise from straight–forward and carefully cued questions. The corresponding level in *Model 1B* had 11 measures in two contexts thus giving what we believe to be a more reliable indication of performance at level 5.

We believe, therefore, that with careful attention to the measures of performance in SoAs as discussed above, we can produce valid and reliable measures of performance in attainment in targets.

4. Conclusions

There are several conclusions to this analysis.

- There are dangers in concentrating on achieving high reliability at the expense of validity. In particular reliable but narrow assessment can have a detrimental backwash effect on the teaching of science. In addition, decisions cannot be based just on statistical analyses but involve a large element of professional judgement.
- Statistics alone do not give all the information we need about the reliability of the SoAs. It is necessary to look at the tasks and assessment guidelines to ensure that they are doing what we want them to do, i.e. to ensure that they are valid.
- In order that pupils are enabled to show the maximum of what they are capable it is necessary to write questions so that they prompt the pupils into giving the responses which are being looked for. The criterion–referenced nature of the decisions which are being made means that the pupils need to be made aware of the criteria which are being used.
- In order that teachers make valid and reliable decisions it is important to ensure that the assessment guidelines match the statement of attainment and that clear exemplars of performance be given.
- Individual responses from pupils have to be seen as performance indicators. They constitute evidence which may be used to make decisions about whether a pupils is performing at a particular level.
- There are different kinds of SoA, and a decision about achieving satisfactory reliability is different for each. For example, a statement of attainment which is concerned

with a single attribute about knowing something as a fact can is quite different from one which concerns several attributes involving understanding.

- In order to get a sufficiently reliable indication of performance in a statement of attainment there should be more than one measure and more than one context.

Acknowledgements and disclaimer

This paper is based upon the work done by the Science Development Team of the Consortium for Assessment and Testing in Schools based at King's College, University of London. The paper could not have been written without the contributions of all the members of the team.

The work was done under contract to the School Examinations and Assessment Council, and I am grateful to the Council for permission to use data from the Final Report of the 1990 Trial (Consortium for Assessment and Testing in Schools, 1990) and to reproduce the *Holiday Centre* material.

The use of this material does not indicate that its content or conclusions represent the policy of the School Examinations and Assessment Council.

References

Consortium for Assessment and Testing in Schools (1990). *CATS Science Key Stage 3 Trial 1990: Final Report.*

DES (1988). *National Curriculum. Task Group on Assessment and Testing: A Report* HMSO.

DES (1989). *Science in the National Curriculum,* HMSO.

DES (1990a). *The Education (National Curriculum) (Assessment Arrangements for English, Mathematics, and Science) order 1990* HMSO.

DES (1990b). *The Education (Individual Pupils' Achievements) (Information) Regulations 1990,* HMSO.

DES (1991) *Science for ages 5 to 16 (1991),* HMSO.

ERIC

Appendix A/1

HOLIDAY CENTRE

School ....

Name ....

Teaching Group ....

Look at the pictures of Jamie and his family on holiday.

❶ Pick out four examples where energy is being used.

Write your examples in the table below. One example has been done for you.

Example	Gets its energy from . .	
Sailing boat	*Wind*	
		13/3a
		13/3a
		13/4a
		13/4b

❷ Explain what will happen to the sailing boat

a if the wind blows harder

______ 13/3a

b if the wind stops blowing

______ 13/3a

❸ Fill in the gaps to show the energy transfers which take place

a when Jamie plays a tape in his personal stereo

______ energy to ______ energy 13/4c

b when Jeni swims across the pool

______ energy to ______ energy 13/4c

❹ Jeni needs energy to pedal her bike. Jeni says,
'The energy I use up riding my bike first came from the sun.'

Jeni is right.

Explain how the energy Jeni uses could have been transferred from the sun, and ends up making her bike move. You will need to include more than one transfer.

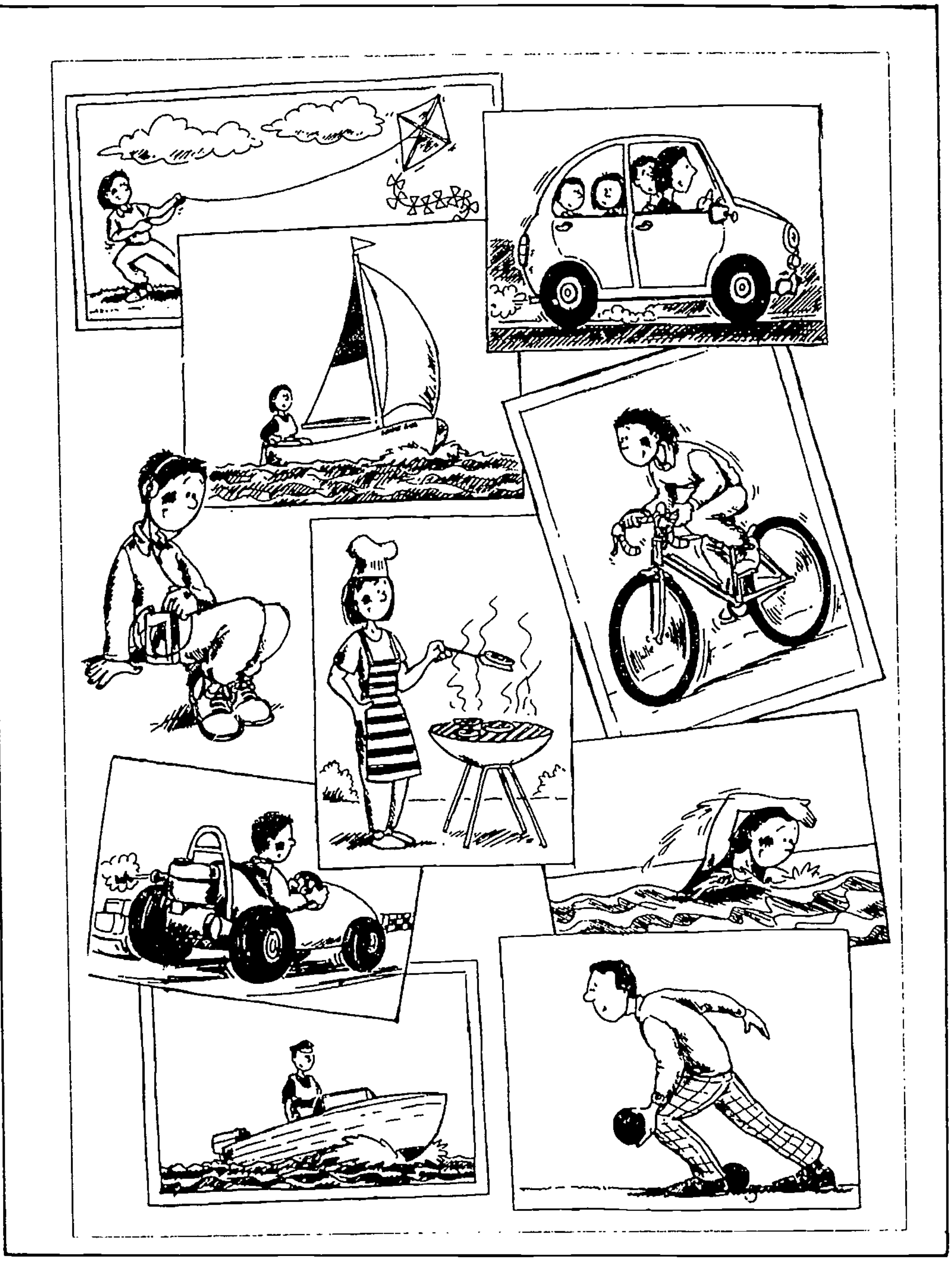
______

______

______ 13/6a

______ 13/6b

Appendix A/2



BEST COPY AVAILABLE

Appendix A/3

Holiday Centre

Teacher assessment guide

Pupil response			evidence of
1		1 "machine" plus correct energy source eg "car" plus "petrol"	13/3a
		more than 1 **Example;**	
OR			
		Examples have entries in **Gets its energy** from column **Examples** don't have to be paired with correct energy source; but all are terms in **Gets its energy** from are "energy words"	13/3a
		includes **Examples** of activities (eg kite flying) with **Gets its energy from** (eg food)	13/4a
		Gets its energy from contains a number of fuels eg charcoal, food/named food, petrol, wood NOT electricity	13/4b
2	a	goes faster	13/3a
	b	slows down/ stops/ doesn't move	13/3a
3	a	potential stored chemical electrical electricity — energy to — sound movement heat	13/4c
	b	chemical stored potential food — energy to — movement heat	13/4c

Appendix A/4

4

Sun to plants	Energy from sun goes to plants plants absorb/change/use the energy to make new plant tissue/grow
Plants to Jeni	Jeni eats plants/ animals eat plants Jeni eats animals
Food energy to movement	Jeni's animal/ plant food releases energy Jeni's muscles use the energy to move/ pedal bike

1 transfer from each box 13/6a

OR

Chain representing above transfers

e.g. Sun → plants → animals/named animals
grass
named plant

↓

Jeni

↓

movement

(arrow heads do not have to be right)

There might also be evidence for part of 3/6b light energy is needed for photosynthesis

Appendix B/1

Attainment target 13: Energy

Knowledge and understanding of science, communications and the applications and implications of science (ATs3-4, 6, 8-11, 13-14)

Pupils should develop their knowledge and understanding of the nature of energy, its transfer and control.

They should develop their knowledge and understanding of the range of energy sources and the issues involved in their exploitation.

LEVEL **STATEMENTS OF ATTAINMENT**

Pupils should:

1
- a • understand that they need food to be active.
- b • be able to describe, by talking or other appropriate means, how food is necessary for life.

2
- a • understand the meaning of 'hot' and 'cold' relative to the temperature of their own bodies.
- b • be able to describe how a toy with a simple mechanism which moves and stores energy works.

3
- a • understand, in qualitative terms, that models and machines need a source of energy in order to work.
- b • know that the temperature is a measure of how hot (or cold) things are
- c • be able to use simple power sources (electric motors, rubber bands) and devices which transfer energy (gears, belts, levers)

4
- a • understand that energy is essential to every aspect of human life and activity
- b • know that there is a range of fuels which can be used to provide energy
- c • understand that energy can be stored, and transferred to and from moving things.
- d • be able to measure temperature using a thermometer
- e • be able to give an account of changes that occur when familiar substances are heated and cooled

5
- a • understand the need for fuel economy and efficiency.
- b • understand the idea of global energy resources and appreciate that these resources are limited.

BEST COPY AVAILABLE

Appendix B/2

LEVEL	STATEMENTS OF ATTAINMENT
6	a • be able to recognise different types of energy source and follow some processes of energy transfer in terms of the principle of conservation of energy.
	b • understand that energy is conserved, but becomes spread around and so is less useful.
	c • be able to explain the distinctive features which make machines, such as pulleys and levers, useful in everyday life.
	d • understand that the Sun is ultimately the major energy source for the Earth
7	a • understand energy transfer by conduction, convection and radiation in solids, liquids and gases and the methods of controlling these transfers, particularly of insulation in domestic and everyday contexts
	b • know that efficiency is a measure of how much energy is transferred in an intended way
	c • be able to evaluate the methods used to reduce energy consumption in the home
8	a • understand that the ultimate result of energy transfers is to change the temperature of the surroundings and that useful energy is dissipated
	b • understand that the use of any energy resource involves both economic and environmental costs, and that such costs may differ in nature and magnitude, depending on the energy source involved.
	c • be able to describe in outline how electricity is generated in power stations from different energy sources, including fossil fuels, nuclear fuels and renewable energy sources
9	a • be able to use the relationships between force, distance, work, energy and time, to describe, explain and compare the functioning of everyday devices
10	a • be able to demonstrate the application of the principle of conservation of energy, and to explain energy transfers in terms of this principle
	b • be able to evaluate the various costs and benefits of different energy sources and appreciate that society needs to take these into account before making appropriate decisions on policy.

SUMMARY OF THE PLENARY DISCUSSION

Jutta Theißen, Universität Dortmund, Germany

In his paper Robert Fairbrother presented facility values (F–values) of questions in standard assessment tasks for Year 9 (14 year old) pupils. The questions were written to assess particular statements of attainment in the English National Curriculum in Science. The answers were marked as either right or wrong, that is, the pupils had either mastered the statement or not. It was shown that the F–values for a statement of attainment varied very much depending on the context of the question and the marking guidelines given to the teacher. Furthermore, it was pointed out that the pupils and their performance were influenced by the way in which the question was written: on the one hand badly formulated questions may cause pupils' misunderstanding of the task resulting in lower F–values, on the other hand cleverly formulated questions may result in very high F–values. This led to the question what F–values meant in the end. The law prescribes the statements of attainment which means that teachers have to assess whether the pupils have fulfilled a particular statement or not. A particular problem is that, while an F–value gives information about the percentage of pupils who got the question right, it is difficult to decide exactly what the correct F–value should be. For Year 9 pupils you expect that the F–value for questions aimed at, say, level 3 (a low level) will be higher than those aimed at level 6 (a high level). This pattern was not always achieved. It was suggested that making the right decisions entailed a mixture of interpreting statistics and using professional judgement about the suitability of the question and its mark scheme.

The advantages and disadvantages of a national curriculum were discussed. A national curriculum gives a common framework which for example makes it easier to move from one school to another, and which enables teachers to communicate with parents. It also means that for the first time all pupils will receive an education in science from age 5 to 16. However, there are considerable difficulties in prescribing appropriate levels of attainment which show progression in learning. It is particularly difficult to get a valid measure of a pupil's level of attainment.

STRATEGIES STUDENTS USE TO SOLVE CHEMISTRY PROBLEMS

Myra L. Halpin, University of North Carolina, USA

Abstract

Previous research studies in chemistry education examined the problems students have in learning chemistry from the viewpoint of the chemical tasks presented to the student rather than from the students' mental strategy (Gabel and Sherwood, 1983). Research literature both at the college and the high school level reports numerous problems preventing students from learning chemistry. Problems identified are: lack of prerequisite skills; unfamiliar units and terms; students; inability to work multiple step problems; too much information given at one time; concepts too abstract for the student; inadequate problem solving skills; lack of understandi ıg the mole concept; poor spatial ability; and lack of cognitive readiness. Gabel (1984) reports that very little research has been conducted to examine, from the students perspective, how high school students solve chemistry problems. Larkin and Rainard's 1984 study suggests that the primary need from research in science education is knowledge that would guide to a better strategy for educating our students to think. They state that we need research to show what students are doing as they struggle with problems or with unit conversions, where they make errors and why.

'o understand the underlying thought processes leading to good problem solving, it is essential to observe in detail the thought processes of individual students (Reif, 1984). Paper and pencil tests are not adequate to ascertain this information because students often do not know what method they used to solve the problem nor why they use that method (Flagerty, 1975). To find out how students solve chemistry problems it is necessary to ask them and to allow them to verbalize what they are thinking, imaging, and processing as they work problems. The presence of an interviewer to probe the student about why they did what they did is essential. Unless educators know the relationship between answers given by students and the thought processes that lead to those answers, there is no accurate measure of the effectiveness of instructors or instruction (Ericsson and Simon, 1984).

The aim of this research was to reveal the students' habits of thinking as important factors responsible for academic success or failure. To do this requires an exploratory research method. Information–processing techniques by Larkin and Rainard (1984) was applied to the protocols collected using the Bogdan and Biklen methodology (Bogdan and Biklen, 1986).

1. Classification of Student Strategies

The first step was to develop a classification scheme of strategies students might use in solving chemistry problems. The scheme was initially developed in a pilot study in which the researcher interviewed twelve high school students as they solved chemistry–like problems (no knowledge of chemistry was needed). These problems varied from simple conversion of ounces to pounds or tons to complex conversions using nonsense words and large decimal numbers. The classification scheme developed in the first pilot study was then used in the second pilot study. The researcher interviewed eight of her own chemistry students as they solved the chemistry problems she had designed for the research study. Modifications to the classification scheme were made as new factors were observed. Tape recorded interviews, problems solutions, and interviewer's field notes constituted the protocols used in the coding for each student to identify the strategies students employed while solving chemistry problems. In the research study, modifications were made during the first eight interviews. No additional changes were made after this time. The Classification of Student Strategies listed in Figure 1 was used to collect the data reported in this paper and from which concludes were drawn for this research study.

Figure 1: Classification of Student Strategies

I. Attacking the problem

- A. Initial Reading
  1. incorrectly — proceeds
  2. number of times
     - a. 1x correctly — proceeds
     - b. reads 2x — proceeds
  3. treatment of problem parts
     - a. reads only first part
     - b. reads entire problem
     - c. chunks parts together
- B. Utilization of equation
  1. always writes if possible
  2. writes if stoichiometry
  3. only if instructed to do so
  4. recognize net ionic
- C. Classifying problem
  1. no classification
  2. classification by principle
  3. classification by general type
- D. Direction in attacking the problem
  1. forward
  2. backward
- E. Analysis
  1. means–end analysis
  2. formula analysis
- F. Images described
  1. pictorial
     - a. 2 dimensional
     - b. 3 dimensional

2. abstract
   a. former problems
   b. equations
   c. systems
3. none

G. Writes down important information

H. Utilizes factor label bar

II. Relevant knowledge

A. Algorithms
1. with understanding
2. without understanding
3. malapropos use of

B. Assumptions made
1. correct
2. incorrect

C. Recognition of extraneous data
1. immediately
2. after mental calculations
3. after trial and error
4. not at all

D. Prerequisite skills
1. has those necessary
2. aware/asks for unknown information
3. familiar words — can't remember
4. "never worked type before"

III. Behavior when obstacle arises

A. rereads problem

B. checks math

C. refers to equation or formula

D. quits

E. asks for help

F. draws a picture

IV. Confidence Level

A. confident with justification

B. confident without justification

C. lacks confidence but successful

D. lacks confidence with justification

V. Problem Types Preference

A. prefers abstract problems

B. prefers concrete problems

VI. Thinking about the problem

A. thinks moles
B. "like a problem I've worked before"
C. equation application
D. real life situation

VII. Method of Checking Problems

A. reread problems and looks at answer
B. math works out — problem must be right
C. checks for logic
D. checks only in the math
E. dimension analysis
F. works backward

2. The Study

The purpose of this research was to ascertain (1) how students solve chemistry problems; (2) to identify indicators or groups of indicators that might be used as predictors of the strategy a student may use; and (3) to determine if some strategies are more successful (produce correct answers) and/or efficient (require fewer steps) than other strategies in solving chemistry problems. Problem–solving in this research is defined as the ability to solve simple and multiple step chemistry problems found in high school chemistry textbooks (Gabel, 1983). The problems used in this study examine the student's understanding of the mole concept as it relates to formulas, equations, and gases. The student's imaging of the three dimensional spect of chemical species was also explored.

Fifty–five chemistry students at North Carolina School of Science and Mathematics in Durham, North Carolina were the target population. All students had the same instructor and the following data were collected on each student:

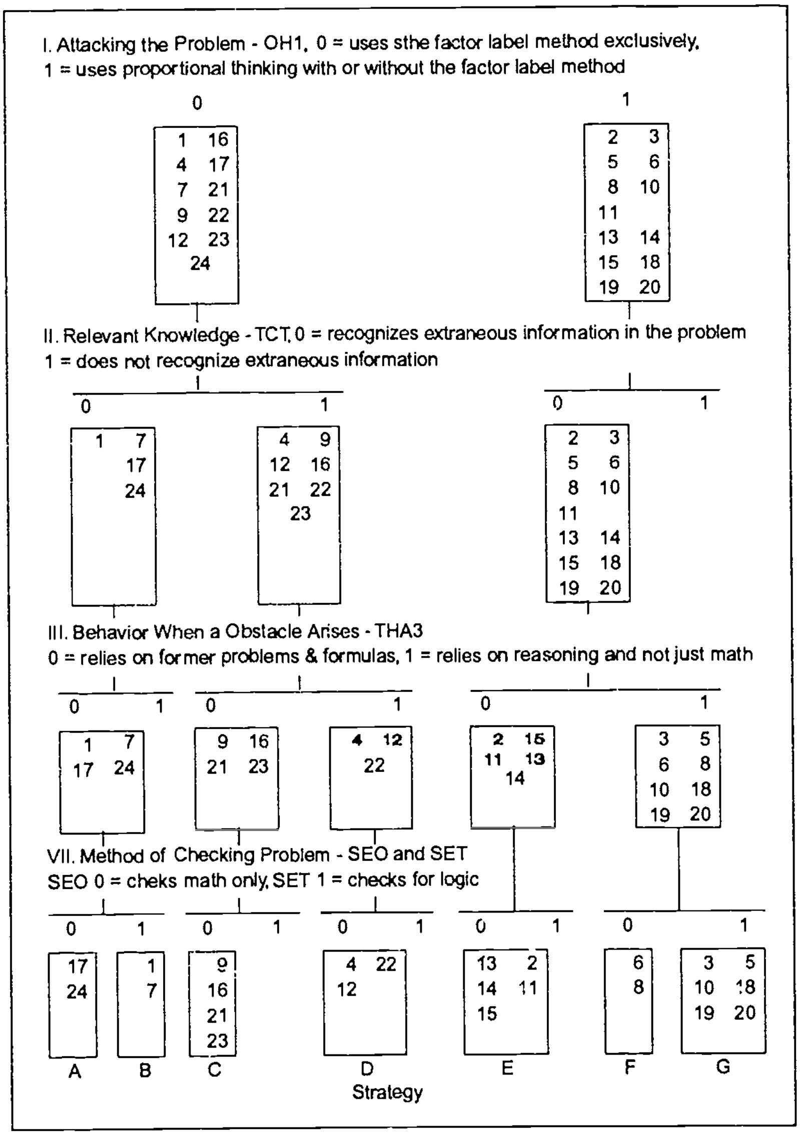
- Gender
- SAT–Verbal
- Sat–Math
- Raven Matrix Score
- Myers–Briggs Type Indicator
- Achievement in Chemistry
- Spatial Ability — Flags and Cubes
- Piagetian Level of Development — Arlin's Test of Formal Reasoning

Stratified sampling techniques were employed to provide a representative range of students for interview. Twenty–four students were selected for ninety minutes, individual interviews in which they were asked to solve twelve chemistry problems. Each student's protocol was coded using the Classification of Student Strategies.

3. Results

The coded information was analyzed using SAS Factor Analysis. Four factors indicated as having high eigen values were used to categorize the students into seven clusters as shown in Figure 2:

Figure 2: Cluster Model Two by Strategies



Each cluster group was then analyzed to determine what other common characteristic the group employed in solving chemistry problems. The common strategies are listed in Figure 3:

Figure 3: Common strategies

Strategy A (students 17 and 24)

- uses the factor label method exclusively
- recognizes extraneous information in the problem
- relies on former problems and formulas
- checks the problem by only checking the math

- reads problem one time then begins
- classifies problems
- uses means–end analysis
- writes down information from the problems
- uses algorithms with understanding
- does not draw pictures
- prefers abstract problems
- writes equations even if not necessary
- does not recognize ionic equations

Strategy B (students 1 and 7)

- uses the factor label method exclusively
- does not recognize extraneous information
- relies on former problems and formulas
- checks for logic not just math

- reads the problem two times before beginning
- does not classify problems
- sees 3–dimensional aspect of chemistry
- sees equations or former problems
- does not use trial and error methods
- does not draw pictures
- prefers concrete problems
- does not think moles
- writes equation but does not recognize ionic species

Strategy C (students 9, 16, 21 and 23)

- uses the factor label method exclusively
- does not recognize extraneous information
- relies on former problems and formulas
- checks only the math

- classifies problems
- writes down information from the problems
- uses algorithms without understanding
- does not draw pictures
- references former problems and equation
- prefers abstract problems
- does not think moles
- writes equations but does not recognize ionic equations

Strategy D (students 4, 12 and)

-- uses factor label method exclusively
- does not recognize extraneous information
- relies on logic and not just math
- checks math and logic

- reads the problem one time then begins
- reads the entire problem before beginning
- classifies problems
- sees equations or former problems
- writes down information from the problem
- uses trial and error methods
- references former problems
- does not draw pictures
- writes equations but does not recognize ionic equations

Strategy E (students 2, 11, 13, 14 and 15)

- uses proportional thinking with or without factor label
- recognizes extraneous information in the problem
- relies on former problems and formulas
- check for logic but math logic not chemistry logic

- reads the entire problems before beginning the problem
- sees 2-dimensional only
- makes charts of information from the problem
- not necessary to reference former problems or equations
- writes equations even if not necessary

Strategy F (students 6 and 8)

- uses proportional thinking with or without factor label
- recognizes extraneous information in the problem
- relies on former problems and formulas
- checks the problem only by math

- uses formula analysis
- sees 2-dimensional only
- makes charts from information in the problem
- writes down information from the problem
- uses algorithms without understanding
- immediately recognized extraneous information
- uses trial and error methods
- does not draw pictures
- no reference to equation or problems
- prefers concrete problems

Strategy G (students 3, 5, 10, 18, 19 and 20)

- uses proportional thinking with or without factor label
- recognizes extraneous information in the problem
- relies on logic to solve the problem

ERIC

- relies on logic to check the problem
- reads the problem one time
- classifies problems
- means–end analysis
- uses algorithms with understanding
- does not rely on former problems to solve the problem

Strategy groups 'A' and 'G' correlate with high SATM and ATFR but differ in that 'G' also correlates with high SATV and 'Thinking' Myers–Briggs types. Strategy group 'F' correlates with low Raven, SATM, SATV, and ATFR scores and all members of the group are female and have 'Sensing' and 'Feeling' Myers–Briggs types. Strategy group 'B' differed from the population only in having low SATV scores and group 'C' in high spatial perception scores. Strategy groups 'D' and 'E' were not significantly different from the research population scores on any of the independent variables.

Each strategy group was compared by success rate (number of problems worked correctly) on the twelve chemistry problems and by the efficiency (number of step taken in getting the solution) with which they solved these problems. Strategy groups 'A' and 'G' are both successful and efficient and strategy group 'F' is unsuccessful and not efficient.

4. Conclusions and Implications for Teaching and Researching

The classification scheme is a useful toll for identifying student strategies used to solve chemistry problems. Two experts, one a qualitative researcher the other a chemist, were asked to code three student protocols using the scheme. There was 91 % agreement between the two experts and the researcher. This scheme could be used by teachers and possibly by students to identify strategies used in the classroom. Science and mathematics educational researchers could modify and use the instrument in their respective fields. The Myers–Briggs Type Indicator also proved to be a useful research toll for this study.

Students need to think about thinking. Metacognition should become a part of chemistry curriculum and instruction. If one knows how they learn best, they may be able to promote their own learning. Three of the twenty–four students volunteerily expressed the benefits of being forced to think about how they think. Teachers need to understand how their students learn best in order to design activities to promote more learning.

During the closure of each interview, the student was asked to recommend changes in the instruction of chemistry that would help them learn more effectively. One strategy group 'A' student, reported "Its fun to watch a precipitation and such — all the colors — but it doesn't really help you on a test that involves math. We should use the class time working more problems." A strategy group 'F' student, reported "I need more demonstrations — work less problems and spend more time on things." Another group 'F' student stated that working simple math problems before the more difficult ones would benefit her learning. From these observation it is apparent that students have different experiences to promote learning.

It is evident from the protocols that students have different experiences in the same classroom. Some students perceive chemistry as a math course because so many of the tests contain mostly math type problems. One distressing finding is that many students who make high grades in chemistry have poor conceptual understanding of chemistry. Since the student is successful using 'plug & chug' algorithms the teacher mistakenly assumes the student has developed a knowledge of chemistry. More emphasis should be placed on conceptual understanding and less on getting mathematical 'right answers'.

This exploratory study is just a beginning. A replicate study using a larger, more representative population of high school chemistry students needs to be conducted before generalizations to a general population are made. The **Classification of Student Strategies** needs further field testing.

References

Anamuah-Mensah, J. (1986). *Cognitive Strategies Used by Chemistry Students to Solve Volumetric Analysis Problems*. Journal of Research in Science Teaching, vol. 23 (9), 759 - 769.

Anamuah-Mensah, J., Erickson, Gaalen & Gaskell, J. (1987). *Development and Validation of a Path Analytic Model of Students' Performance in Chemistry*. Journal of Research in Science Teaching, vol. 24 (8), 723 - 738.

Bender, D. S. & Milakofsky, L. (1982). *College Chemistry and Piaget: The Relationship of Aptitude and Achievement Measures*. Journal of Research in Science Teaching, vol. 19 (3), 205 - 216.

Bloom, B. & Broder, L. J. (1950). *Problem-solving Processes of College Students*. Chicago: The University of Chicago Press.

Bodner, G. M. (1986). *Constructivism: A Theory of Knowledge*. Journal of Chemical Education, vol 63 (10), 873 - 878.

Bodner, G. M. (1987). *The Role of Algorithms in Teaching Problem Solving*. Journal of Chemical Education, vol. 64 (6), 513 - 514.

Bogdan, R. & Biklen, S. (1982). *Qualitative Research for Education*. Boston: Allyn & Bacon.

Bunce, D. M. and Heikkinen, H. (1986). *The Effects of an Explicit Problem-Solving Approach on Mathematical Chemistry Achievement*. Journal of research in Science Teaching, vol. 23 (1), 11 - 20.

Camacho, Moisers, Ron Good (1989). *Problem Solving and Chemical Equilibrium*

COMMENTARY OF THE DISCUSSANT

Elke Sumfleth, Universität Essen, Germany

This paper of discussion highlights some facts of the presentation by M. Halpin being important above all. The statement — that students and teachers need to know how they respectively their students learn best — will meet with unanimous approval. I agree with you that "students with different learning styles need different experiences to promote learning", but I have doubts that the students' statements you quoted depend on their learning styles. These statements depend much more on the content treated and on the grade of achievement the students have already obtained:

Group 'A' student: *"Its fun to watch a precipitation and such — all the colours — but it doesn't really help you on a test that involves math. We should use the class time working more problems."*

Group 'F' student: *"I need more demonstrations — work less problems and spend more time on things."*

Group 'F' student: *"... that working simple math problems before the more difficult ones would benefit (my) learning."*

The first and the third statement differ only gradually. A beginner who is not able to solve complex problems has to start with simple ones and later on — like the 'A'–student — he wants to exercise more difficult ones. Regarding the first and the second statement there is no real contradiction. The 'F'–student does not demand for a demonstration or a 'thing' like 'precipitation' because he really might not see any relationship between a precipitation reaction and stoichiometric problems. A statement like the first one seems to be an indicator on the function of experiments in class rather than an indicator on problem solving strategies of students.

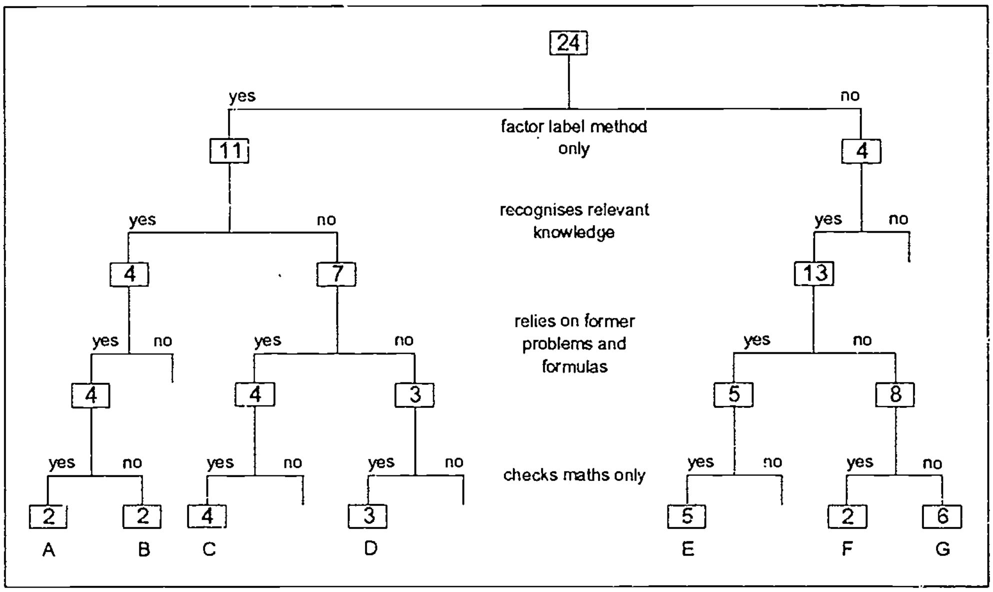
The fact that students perceive chemistry as a math course is particularly pronounced in this context. The finding that many students make high grades in chemistry without any conceptual understanding corresponds to our results (Sumfleth, 1988a and 1988b); are able to solve stoichiometric problems by using algorithms without understanding the theoretical context. And Yarroch considers 1985 (Yarroch, 1985) "... the majority of the students were not able to demonstrate that they knew anything more about chemical equation balancing than the mechanical manipulation of symbols." All students had been selected for that study because of their performances in chemistry which were evaluated well by their teachers. The question arises what is the sense of such problems.

But now back to the problem solving strategies the students used in this paper. The students of two groups 'A' and 'G' are successful and work efficient, the students of group 'F' fail. Which are the common strategies, which are different? See the figure on the next page.

Comparing the four factors used for categorising the students Group G and F only differ in one factor, namely the method of checking the problem: Group F checks maths only, Group G checks for logic. This might be one reason for success but the other successful group A checks for maths only too. On the other side group A and G differ in all factors except that one concerning relevant knowledge. Therefore it must be of great importance for successful problem solving to recognise extraneous information in the problem, but those who fail completely, recognise these information, too.

It is possible that the success of group A is a special feature of solving stoichiometric problems. As discussed above, group A might be successful because it is sufficient to use factor label methods exclusively, to rely on former problems and formulas and to check math only if the algorithm used is correct. But why does group B come off worse than A regarding that this group checks the problem for logic? Checking for logic must be a handicap by using algorithms only whereas only checking math is not sufficient by not using algorithms concerning group F (Table 1 on the next page).

Each cluster group was then analysed to determine what other characteristics the group showed in solving chemistry problems. First of all the number of common characteristics differs group by group. There might be a correlation to the number of students belonging to each group. The number of characteristics decreases with an increasing number of students. Most of the characteristics of group G you can find again among those of group A, but none you'll find among those of group F.



These last results are in agreement with the successful/unsuccessful and efficient/inefficient characteristics.

Group A	Group G	Group F
Reads problem one time then begins;	Reads the problem one time;	Uses formula analysis;
classifies problems;	classifies problems;	sees 2-dimensions only;
uses means-end analysis;	means-end analysis;	makes charts from information in the problem;
writes down information from the problems;	uses algorithms with understanding;	writes down information from the problem;
uses algorithms with understanding;	does not rely on former problems to solve the problem.	uses algorithms without understanding;
does not draw pictures;		immediately recognised extraneous information;
prefers abstract problems;		uses trial and error methods;
writes equations even if not necessary;		does not draw pictures;
does not recognise ionic equations.		no reference to equation or problems;
		prefers concrete problems.

ERIC Full Text Provided by ERIC

My question now: How do you explain the great differences between A and G and small ones between G and F concerning the main factors and in contrast great conformity in special characteristics regarding A and G?

I would like to add a totally different question concerning my own research interests: Are pictorial images of any importance? (Sumfleth, 1991)

Finally, it must be of great interest to spread this investigation in two directions:

(a) to extend the number of participating students
(b) to change the chemistry contents.

References

Halpin, M. (1992). *Strategies students use to solve chemistry problems*, paper presented at the international seminar "Empirical research in chemistry and physics education", Dortmund.

Sumfleth, E. (1988): *Lehr- und Lernprozesse im Chemieunterricht*, Frankfurt/Main, Bern, New York and Paris: Peter Lang

Sumfleth, E. (1988). *Knowledge of terms and problem-solving in chemistry*, International Journal of Science Education vol. 10 (1), 45 - 60

Sumfleth, E. and Körner, H.-D. (1991). *Mentale Repräsentationen — ein lernpsychologischer Konstrukt mit Bedeutung für die Chemiedidaktik?*, Der mathematische und naturwissenschaftliche Unterricht vol. 44 (8), 458 - 463

Yarroch, W. L. (1985). *Student understanding of chemical equation balancing*, Journal of Research in Science Teaching vol. 22 (5), 449 - 459

SUMMARY OF THE PLENARY DISCUSSION

Holger Eybe, Universität Dortmund, Germany

The statements of students from different strategy groups ("It's fun to watch a precipitation...", etc.) were merely given to provide a better understanding of the groups. They do not indicate what strategies the students used but rather how they perceive themselves to learn best. Therefore, these responses were not part of the coding.

As for the differences between the groups A and G, the main difference is *how* they process the information. All groups can be successful in solving problems, also group F, for example. Women in science have the same potential as men, perhaps here the problem is a question of structuring the curriculum.

The researcher omitted the confidence level from the coding because she felt it was too subjective. It was discussed whether or not the confidence level should be part of the coding. Perhaps this measurement is too subjective. On the other hand, students could be asked how confident they are, it could be measured on a scale from 1 to 10. It can be observed that females are not as confident as males, but just as successful in their work.

It was suggested to investigate what students say to each other while solving problems. Different methods could be used, for example taping students discussing the problem as they work or interviewing them afterwards so that they can reflect on what they have learned. It is always difficult to obtain answers that can be used to identify students' strategies unless probing questions are asked.

Quite a large number of statistical tests were used for the investigation. Thus, there is the danger of multiple testing. The statistical tests were not supposed to yield quantitative or significant results nor to fit in with other models. It was merely tried to obtain some information about what ways of thinking, strategies and conceptions there are by using an easy to administer test. Perhaps a descriptive analysis might be the best method. From a teacher's point of view it can be a valuable experience for both the teacher and the student to try and find out what the student knows and thinks after four or five months of teaching. Students enjoy talking about how they think and what they think.

As for some typical examples of strategies students from different groups use: there is a strategy group A person who always has very good marks. He says you cannot use logic with gas mole problems. However, he knows how to get the correct answer by using an algorithm. There seems to be a tendency towards teaching one right answer to everything. But the researcher feels that after learning algorithms there has to be logic as well. Strategy group G students need to see the whole picture first, then they consider the smaller pieces. If they only have some pieces they always feel that something is missing. There are also approaches that do not seem sensible to the teacher, but that are successfully used by individual students.

Apparently it is easier to learn something or to solve problems if there is a context. For example, people who want to find out how much they have earned in 1.5 hours usually get along well with using fractions in calculations. Also, a waiter can memorise more orders than a normal person. Everybody should have the time to develop concepts in a structure that makes sense to them.

THE EDUCATIONAL STRUCTURE OF ORGANIC SYNTHESIS

Hanno van Keulen, Universiteit te Utrecht, The Netherlands

1. Teaching organic synthesis: why and how?

In this lecture I will focus on the main theme of my research project, which is: "The Didactical Structure of Organic Synthesis". The basic research question is: How can I teach students to plan, design, perform, and evaluate a synthesis procedure in organic chemistry? I use the term 'didactical structure' instead of the more familiar 'educational structure' to indicate a set of educational activities, such as tasks and questions, including the chemical content and sequence, which is *based on empirical research*. I will return to the research method in the next part of this lecture.

But I will start with saying a few words on the goals and contents of chemical education. In this, I will restrict myself to *organic synthesis at the university level*. At this level, the main educational goal is, or should be, 'preparing students for carrying out scientific research'. In order to prepare students for this goal, we should know what it is when someone is able to carry out scientific research. A researcher needs two different sets of things. In the first place, he or she needs professional qualities. In the second place, there should be a scientific problem and a way to investigate. I call this second aspect the *context of scientific inquiry*.

The professional qualities, in general, are: *knowledge* of the substantive structure of organic chemistry, in this case knowledge of compounds, molecular formulae, reactions, mechanisms and principles; practical *skills* with regard to laboratory techniques; *problem solving* capability; and last but not least *creativity*. But, can we teach these qualities?

To start with creativity, it is a fact that some students come up with brilliant solutions and ideas, whereas others do not. But, at least to me, it remains quite unclear how to *enhance* this capacity with educational means. I think the only thing we can do is to leave enough room for students to follow their own creative ideas. I will not consider teaching creativity in this lecture.

Traditionally, education in organic chemistry strongly emphasises the knowledge of facts and techniques. Factual knowledge is taught with the help of lecture courses and textbooks. Laboratory techniques and practical skills are dealt with in lab courses.

2. Problem solving in organic chemistry

This leaves us with problem solving. Problem solving can be defined as that what you do when you do not know what to do. There are three main ways to solve a problem: by creativity, by luck and/or by rational thinking. Once again I will not deal with the creative insights. Solutions by luck and serendipity are equally hard to promote by means of education, which leaves us with rational thinking. So this should be the major goal for education in organic synthesis.

But what exactly does it mean to be *rational* in synthesis? In my opinion, it has to do with the ability to generate new scientific facts. And this depends on the ability to recognise regularities and diversities among observations, the ability to recognise problems, understand experimental methods, organise and interpret data, understand the relation of facts to the solution of problems, plan experiments and make generalisations and assumptions. I call this *the rationale of organic synthesis*. Rational thinking in synthesis is applying the rationale, which is not identical with knowing a lot of facts. So we cannot depend on the lecture course to promote rational thinking.

Hence, in my research project, I cannot restrict myself to the 'paper' part of problem solving, e. g. the design on paper of a reaction route from starting materials to products. I think that the laboratory part of organic synthesis should receive as much attention as the theoretical part. It is my experience that many students who can produce a correct reaction formula do not have the slightest idea how to actually carry out this reaction. Knowing things 'in theory' often is a euphemism for not knowing things at all. The origin of this problem might be a neglect of what I call the *context* of science. I will now explain some of my educational paradigms, in which I need to elaborate on the meaning of this word 'context', and also on the word 'concept'.

3. What is a 'concept'?

There seems to be general confusion concerning these words. First, 'concept'? Of course, many different descriptions exist. I will simplify matters somewhat by stating that the word can be used in an *objectivistic* and a *subjectivistic* sense. In the objectivistic sense, concepts are the means by which a *discipline*, such as chemistry, structures its facts. Thus, there is an 'element' concept, a 'substance' concept, a 'reaction' concept, and, in organic synthesis, there is a 'Lewis acid' concept, a 'nucleophile' concept, and so on. You can find descriptions and definitions of these objectivistic concepts in the textbooks.

In the subjectivistic sense, it has to do with the way *individuals* structure their knowledge. So *I* have an 'element' concept, a 'substance' concept, etc. I think it is better to call this '*conception*' instead of 'concept'. However, when I use such words as 'reaction', or 'acid', I do not only intend the textbook definition, but I also have in the back of my mind all my *tacit knowledge*. This tacit knowledge includes associations and memories of phenomena, theories, definitions, actions, scenarios, skills, and previous situations in which the word was encountered. These conceptions are not objects which can be defined objectively and completely. The reason that successful communication between chemists is possible is because their words trigger the same associations. But this does not necessarily happen to a novice, who has only learned the definition: the objective part of the conception.

What is meant when someone uses a concept depends not only on the objective sense, but also on the *context* in which it is used. For example, the energy concept in thermodynamics is radically different from the energy concept in the context of sporting. Now this has some implications for educational programmes based on the popular 'conceptual

ERIC

change' ideas. Pupils come to the science class with conceptions of 'the world' which are often quite different from those of science, it is said. Such statements reflect a predominant feeling that science deals with 'nature', and, since we are part of 'nature', scientific knowledge should also hold true for our daily life. I think this assumption is too bold. In most cases, a scientific concept has no real equivalents in the life-world. Science *reduces* the natural world in such a way that all variables can be controlled. But the resulting concepts have no existence in the unreduced life-world. Consider for example the famous case of the force concept in mechanics. A constant velocity means that the net forces are zero. But do you apply such a concept in your daily life, when you are driving your car? No, instead you use a life-world conception of force which is well-adopted to its function in the life-world context. A scientific concept has only meaning in a relevant, that is scientific, context. To use a scientific concept successfully, you have to have knowledge of the context. This is what I intend with the word 'conception': part of it is factual, and part of it is tacit, personal knowledge, acquired through experience. If we want that students acquire these conceptions, we should create the 'right' context for this process. I propose to call this *'conception development'*.

4. A context for conception development

The job of the curriculum designer is to design such a context in which students can experience scientific phenomena and subsequently construct useful scientific conceptions. Students have to be brought in situations in which such a conception development is possible. It is my opinion that the tacit knowledge part of conceptions cannot be communicated through verbal instruction. This implies that conceptions can only be learned completely through direct (which often means: practical) experience, which is the main reason why we cannot do without laboratory courses. When novices are confronted with unfamiliar natural phenomena, they observe, reason and interpret in a highly idiosyncratic way. They cannot be told just to construct the 'right' conception. Consequently, the nature of the experiences in education is very important. I would add another point: it is important that these experiences are made explicit. In a scientific study we cannot be satisfied with tacit knowledge alone. It is through language that scientists communicate, so we should strive to find words for our experiences. This should be reflected in two ways. In the first place, education should give students ample opportunity to discuss matters. In the second place, educational researchers should try to make explicit as much of the tacit component as is possible.

5. Simulation of scientific research

This analysis leads me to the conviction that scientific organic synthesis should be taught in a scientific context, because this is the only way in which student conceptions will converge with the experts conceptions. My proposal for this context is: a *simulation of scientific research*. The purpose of scientific research is to generate new factual and theoretical knowledge. I suggest to simulate this process. However, I do not propose some sort of discovery learning. Students should not be left to themselves, but should be guided by a didactical structure in which questions, tasks and lab experiences can be prescribed.

So now we have an objective, i.e. 'understanding the rationale of organic synthesis', and we have an outline of the context. The next questions to be answered concern the content and structure: which concepts are used in this rationale; and: how can they be taught? For this, I will consider the existing situation first.

Education in organic chemistry traditionally consists of a series of lectures in which factual knowledge is covered, and a laboratory course on the practical aspects and techniques. The lectures and textbooks provide the body of substantive knowledge, the facts. They classify compounds and reactions in relation to their chemical and physical characteristics. In textbooks, reactions are described in a general way in which the molecular formulae are the key features. But, reaction conditions often remain unspecified and activities such as isolation, purification and characterisation are not mentioned in most cases. Since the aim of the textbooks is to provide an overview of the *results* of research, the original research questions, problems and methods have disappeared from the text. This leads to a presentation which seems somewhat misleading: it appears as if the argumentation preceded the discovery of the reaction, whereas in fact this argumentation was provided with hindsight. The established facts are being transferred to students, but not in a scientific context. This implies that students will not be capable to transfer this knowledge to the field of scientific inquiry without substantial help.

However, this need not be a problem. There is nothing wrong with having student acquire factual knowledge of the subject matter. This is a necessary and non–tacit part of conceptions. It just depends on the way we deal with the missing part, the tacit component. And here we encounter the real difficulties.

In the traditional lab courses, the main objectives are *learning laboratory techniques* and *illustrating textbooks reactions*. Syntheses are ordinarily performed on the basis of a 'recipe': a complete description of all laboratory actions which have to be carried out to obtain a satisfying yield and purity. However, very little attention is paid to the rationale behind the prescriptions. As a result students often have no idea why the procedure works so well. Worse, they even do not need to know what the rationale is in order to get good chemical results. Here we confront a major problem: the cookbook problem. The fault of many organic experiments is that there are no questions asked and no thinking done. There are only instructions given to allow students to obtain products. In general, manual skills and laboratory techniques are taught efficiently in this way, but rational thinking is not. Although many prefaces of laboratory manuals contain impressive remarks on 'learning the scientific method', their major objective is teaching techniques and illustrating reaction types. The assumption that students who obtain satisfying chemical results also have learned something on doing research is a mistaken inference.

In conclusion, this short analysis stimulates me to solve this problem by simulating scientific research.

6. Empirical method

This brings me to the questions concerning the development of educational material that will help students learn to understand this rationale. Now it is time for the empirical

ERIC

part. I need a suitable educational research method to develop this material and to investigate its effectiveness.

Basically, my method is a cyclic process of evaluating, developing and executing. In the first cycle the evaluation is a search for the right goals and contents for teaching organic chemistry. Such a search for objectives and contents results in a tentative idea of what we want to teach. These ideas typically arise from teaching experience, intuition and creativity. On the basis of these ideas a first educational design is developed. Before carrying out this design I try to predict what will happen: what exactly will students do and say. Then the design is put into practice. The educational process is followed closely and is carefully analysed, often with the aid of protocols of conversations and discussions between students and the teacher. The analysis provides insights in the way both students and teachers experience and deal with the subject matter. This leads to a second cycle, with a new reflection on goals and contents, a new design, new predictions on what students and teachers will do, and new observations and protocols. Eventually the resulting educational structure should lead to the educational results in a predictable way. At our department, when this is the case, we call this a 'didactical structure'.

What I try to do is to develop criteria on the basis of which successful teaching material can be constructed. This approach is in essence qualitative. I am not comparing different experimental approaches to reach the same objectives, but these themselves change as a result of the evaluation cycles. In the first cycle, my objectives were vague and corresponded in many ways with the traditional goals. But at present, my objectives concern the rationale of organic chemistry which implies a shift from the objectives of the traditional lab course. So, from a methodological point of view, it is not possible to compare these different approaches. That is why I try to convince you with qualitative arguments and not with statistical evidence. Another reason for the absence of statistical material in my research is that so far my interest is in *developing*, not *testing*, concepts, and this, of course, predominantly is a qualitative enterprise.

7. Esters

After this theoretical digression I will now give an outline of an experiment which simulates research. This experiment is called "Esters". Now making esters in itself is quite unproblematic when you follow a recipe. But, when simulating research, there cannot be a recipe. What I am aiming at is not that students make esters, but that they to go through the process of posing and investigating research questions. Now students are not yet professional researchers. During high school, they have acquired some factual knowledge, but they still lack the sophistication of the expert. You cannot expect freshmen to pose scientific questions just out of the blue sky. So what I do is to present them with a problem that I think they can investigate on the basis of their existing knowledge.

The starting question of the first experiment is "How would you make an ester?". Esterification is a topic which in The Netherlands is treated in secondary education. That is to say, students have to learn the reaction and the mechanism as described in a textbook, but most of them never actually perform an esterification. Hence, I can expect that the students will know the general ester-structure, and will be able to recall the reaction formula. Now they think they can make the stuff, which means that we have something like a *problem* (how to make an ester) and a *hypothesis*, in this case: just put an alcohol and a carboxylic acid together. That is what they have learned, and that is what they propose to do. A teaching assistant leads the discussions, but does not give an indication of what is right or wrong. Remember, we are trying to simulate research, and so nobody knows 'the real answer'.

Students then carry out the reaction. Now what I want is not in the first place that students make esters, but that they construct useful conceptions. For this, they need to have more than one experience. You cannot categorise and conceptualise on the basis of a single phenomenon. You have to compare to see regularities and differences. That is why I work with groups of students (groups of eight, to be precise) who carry out slightly different reactions. Comparison and group discussion helps to promote conception development.

Students also carry out additional tasks. These are designed to produce experiences which can be utilised to solve the problems the moment students become aware of them. Some of these problems, such as how to handle the apparatus or how to run the laboratory equipment are solved with the help of a teaching assistant. This is comparable with the traditional laboratory course. But now they also meet with 'real' problems of a more scientific kind. For instance, when you put together the two starting compounds of an esterification reaction, nothing seems to happen. Both alcohol and carboxylic acid are colourless liquids. The mixture remains colourless, and, to the confusion of the students, the characteristic sweet ester smell does not develop immediately. Only after some time has elapsed do some of the reaction mixtures seem to smell sweet to some of the students. So the question is: has a reaction taken place, or not? If not, why not? And, what next? These kind of problems are not solved by the teaching assistant. Instead, the teaching assistant organises a discussion, in which students have to report and discuss their experimental findings. This discussion should result in theorizing on the basis of these experiences, and in new hypotheses which can be experimentally tested. The teaching assistant leads the discussion, and gives technical advice concerning how to do the things students propose. The general outline of this experiment is thus an alternation of discussions in which observations and assumptions are discussed, and experimental testing of the assumptions.

This process eventually can lead to success, not only from a chemical but also from an educational point of view. The students are able to devise a procedure which more or less works. But more important, they develop useful conceptions along the way. This process is carefully recorded: all discussions are (audio) taped and the lab activities are observed. This gives me, as a researcher, the essential information on their conception development. I see it as my objective as educational researcher to state these conceptions explicitly, this is, to bring as much as is possible out of the realm of tacit knowledge.

8. Examples

I will now proceed with three examples of this process. The first example concerns the *reaction mechanism concept*. In the Ester–experiment, two different carboxylic acids are applied: formic acid and acetic acid. Formic acid reacts faster. Since it is the stronger acid, this leads students to the hypothesis that acid strength has something to do with the reaction. This provokes interest into the mechanism of the reaction. What have acids to do with this reaction? Reflection on this observation leads students to the logical assumption to use sulphuric acid as a catalyst. But this is only the beginning. Sulphuric acid is a Brønsted acid: it donates a proton to the carboxylic acid. This is the first step of the mechanism. But, in order to understand the next steps they have to develop a Lewis–acid concept: alcohol acts as a Lewis acid towards the central carbon acid of the carboxylic acid. By carefully analysing the discussions between students, I can observe the influence of the laboratory experiences and the subsequent discussions on their language. Some students manage to make the leap from the Brønsted theory to the Lewis theory, which can be concluded from their sudden use of the Lewis terminology. Words like nucleophile and electronegativity play an essential role in this theory, whereas they do not in the Brønsted theory.

The second example deals with the *reaction spectrum concept*. Not all reactions follow the ideal pattern A plus B gives C. There are many other possibilities, such as equilibrium reactions, side reactions, and subsequent reactions. In the case of esterification, we are dealing with an equilibrium reaction. That implies that when you start the formation with equimolecular amounts of alcohol and carboxylic acid, these will still be present when the reaction has come to an end. Gas chromatography makes this obvious. Those students who understand this can see the consequences for synthesis: one of the starting reagents should be present in excess, or one of the products should be removed from the reaction mixture. This will drive the reaction to higher ester yields. Group discussions stimulate this understanding: all students can refer to the same experiences, have the same problem, and so are interested in each others' solutions.

Knowledge of the actual reaction pattern often has important consequences for the whole synthesis procedure. This brings me to my last example, the *concept of synthesis planning*. By this I mean the relationships between the stages in a synthesis procedure. A typical synthesis consists of formation, working up, purification, and characterisation. In the ester synthesis, students find out in the characterisation stage that alcohol and acid still are present. This was due to the fact that the reaction is an equilibrium reaction. Now they immediately see the need for a purification to be carried out. But, whatever they try, they cannot get rid of the alcohol, because this has almost the same physical properties as the desired product, the ester. During the discussions, these observations are combined into the proposal to start the reaction not with equimolecular amounts, but with an excess of the carboxylic acid. With the help of the gas chromatograms they can even estimate the equilibrium constant and calculate the relative amounts in order to obtain 99 % purity.

These three concepts, reaction mechanism, reaction spectrum, and synthesis planning, provide a rationale for doing organic synthesis. They can be constructed into the corresponding conceptions by students in an educational context which simulates research. In this first experiment, the conceptions of course do not come to their full sophistication. But this study provides me with criteria to develop further laboratory experiments in this context. That is what we are working on at the moment.

References

Boud, D., Dunn, J. & Hegarty-Hazel, E. (1986). *Teaching in Laboratories*. Guildford: SRHE & NFER-NELSON.

Elzenga, H. E (1991). *Kwaliteit van Kwantiteit*. Utrecht: CD-b Press.

Fife, W. K. (1968). *The Organic Chemistry Laboratory*. Journal of Chemical Education, vol. 45 (6), 416 – 418.

Hodson, D. (1991). *Practical Work in Science: Time for a Reappraisal*. St. Sci. Educ., vol. 19, 175 – 184.

Hodson, D. (1992). *Redefining and Reorienting Practical Work in School Science*. School Sci. Rev. vol. 73 (2), 65 – 78.

Jager, H. de. (1985). *Leren Synthetiseren — Symbiose van Ambacht en Wetenschap*. Utrecht: Rijksuniversiteit Utrecht (dissertation).

Jager, H. de, & Verdonk, A. H. (1986). *Synthetiseren: Imiteren of Experimenteren?* Vakdidactiek en Informatietechnologie in Curriculumontwikkeling. Ed. J. S. Brinke, H. P. Hooymayers and G. Kanselaar. Lisse: Swets & Zeitlinger B. V.. 89 – 100.

Kandel, M. & Ikan, R. (1989). *Provocative Replies: Two Organic Chemists Look at Organic Chemistry Lab*. Journal of Chemical Education, vol. 66 (4), 322 – 324.

Kirschner, P. A. (1992). *Epistemology, Practical Work and Academic Skills in Science Education*. Scie. & Educ. vol. 1 (3), 273 – 299.

Mulder, T. & Verdonk, A. H. (1984). *A Behavioral Analysis of the Laboratory Learning Process*. J. Chem. Educ. vol. 61 (5), 451 – 453.

Pickering, M. (1985). *Lab is a Puzzle, not an Illustration*. Journal of Chemical Education, vol. 62 (10), 874 – 875.

Pickering, M. (1989). *Choosing to Cookbook*. Journal of Chemical Education, vol. 66 (10), 845 – 846.

Potter, N. H. & McGrath, T. F. (1989). *Getting Away from the Cookbook in the Organic Laboratory*. Journal of Chemical Education, vol. 66 (8), 666 – 667.

Polanyi, M. (1962). *Personal Knowledge*. 2nd ed.. Chicago: University of Chicago Press.

Tobin, K. (1990). *Research on Science Laboratory Activities: in Pursuit of Better Questions and Answers to Improve Learning*. School Science and Mathematics, vol. 90 (5), 403 – 419.

Venkatachelam, C. & Rudolph. R. W. (1974). *Cookbook versus Creative Chemistry*. Journal of Chemical Education. vol. 51 (7), 479 – 482.

Wade, L. G. (1979). *Chemistry without a Cookbook*. Journal of Chemical Education, vol. 56 (12), 825 – 826.

Wester, F. (1987). *Strategieën voor Kwalitatief Onderzoek*. Muiderberg: Couthinho.

ERIC

COMMENTARY OF THE DISCUSSANT

Kerstin Prokoph, Pädagogische Hochschule Halle–Köthen, Germany

The problem of studying organic chemistry at laboratories on a problem–oriented way, considered by Mr van Keulen in his lecture, could be extended in three directions:

(1) concerning chemical lab courses in general,
(2) concerning teaching chemistry in general,
(3) concerning all students and pupils, who want or have to study chemistry.

Let me discuss the topic of Mr van Keulen in this extended form. I want to start with the third point. In my opinion it is a precondition to know who wants to study chemistry and why. This determines all further contents and methods of teaching. There are different goals in teaching later scientists, teachers, lab–workers, pupils etc. But problem–oriented lab work can be done in all directions of chemical training.

Because I am working at a teachers' training college, I want to make some remarks about problem–solving related to the lab courses for teacher–students of chemistry. For teacher–students there are two fundamental kinds of lab courses:

(1) professional chemical lab courses,
(2) didactical chemical lab courses.

Of course, the two kinds of lab courses content also problem–oriented stages. In this connection I would like to remark, that not the experiment itself leads to problem–oriented work. Also 'cookbook'–experiments can be problem–oriented if the instructor leads them in a suitable way. This can be realised, for example, by interviews during the lab course, but also in preparation to the course at seminars, lectures etc.

Professional and didactical problem–oriented lab work have different goals. In the professional fields of chemistry the students have to develop their skills to solve problems in research. In the didactical field they have to develop their skills related to problem–solving in two ways:

(1) They have to plan and develop experiments for lessons of chemistry. (For example, under the conditions of school life it is necessary to construct cheap simple equipment.)

(2) They have to train leading problem–oriented lab work of pupils.

Although I support problem–oriented learning, I agree with Kandel (1989, p. 322): "... learning problem–solving and 'cooking' together — it's not an either/or situation." I am convinced that it will be necessary in future, too, to obtain fundamental chemical knowledge in theory as well as in the laboratory with traditional methods. I think there are limits, especially in regard with the organisation of problem–oriented lab courses. Problem–solving requires individual solutions. But how to teach a group of individualists? Another fact: My experience in leading problem–oriented learning let me be afraid concerning the motivation and the differences in knowledge of students and pupils — not in relation to problem–oriented learning, but in the professional questions of chemistry. How can be lead problem–oriented learning, if the instructor has to spend his time with motivating and explaining fundamental theoretical connections? I think it is a question of time, staff and money to be successful in leading a problem–oriented learning process.

But only without forgetting proved traditional methods chemistry can be taught successfully.

References

Kandel, M. & Ikan, R. (1989). *Provocative Replies: Two Organic Chemists Look at Organic Chemistry Lab*. Journal of Chemical Education, vol. 66 (4), 322 – 324.

Pickering, M. (1985). *Lab is a Puzzle, not an Illustration*. Journal of Chemical Education, vol. 62 (10), 874 – 875.

Pickering, M. (1989). *Choosing to Cookbook. Journal of Chemical Education*, vol. 66 (10), 845 – 846.

SUMMARY OF THE PLENARY DISCUSSION

Petra Beuker, Universität Dortmund, Germany

Hanno van Keulen claimed that one of the major problems in his work is training teachers. It focuses on didactical problems and the question of teaching them how to teach in an appropriate way. Van Keulen has observed that teachers often disturb students' discussions by changing the subject because they, the teachers, are bored. It has to be considered that teachers also have a long career in problem–solving in science. They should be taught that it is much more effective to learn in a group. The advantage of teamwork is that the students can profit from one another instead of competing against each other. Teamwork, instead of hopping from technique to technique in the traditional way, makes teaching more efficient. Although it might take more time. Cookbook experiments can be helpful if teachers are prepared to teach them in a suitable way. A lot of useless laboratory work should be skipped. This is also a way of saving money.

There is no recipe for which parts to teach in cookbook manner. The teacher has to find out empirically. Explaining the equipment in the laboratory should be taught in the cookbook manner. However, problems that appeal to their prior knowledge, like e. g. making an ester, should be solved by themselves. The lack of direction should not cause frustration among the students. Maybe the teacher should start with a lot of directing and then let them do more and more themselves. The problem should always be manageable. Some problems may seem manageable for the teacher, though the students have their difficulties. Therefore the problem should be thought over and changed again and again in a cyclic approach. It always depends on the students whether they like this type of teaching or not, but one can say that there will always be a few who will not like it. The time for the students' discussion should be limited, because if it lasts too long they will be frustrated if they do not come to a conclusion. It sometimes happens that the teacher gets tired after more than one hour of talking about chemistry. But after 40 – 50 minutes they should come to a conclusion. Students want to discuss, but they do not want to do things. Most of them are afraid of taking the responsibility for doing an experiment. Therefore they try to avoid it by discussing. Because of that the teacher has to give a certain amount of time for discussion. After that he should give them tasks and

ERIC

clear orders. This makes laboratory activities more interesting.

It is easier to teach this way at a university level than in secondary schools. Hanno van Keulen questioned the aims of teaching chemistry in secondary schools. There is no use of understanding chemistry if they do not need it for their everyday life. E. g. the students do not have to know about chemistry to read and understand a newspaper.

However, the reason for teaching it in secondary schools is that chemistry is an intellectually challenging culture, e. g. the fact that only the functional groups react with each other when two things are mixed together. Students have different misconceptions, but if they have grasped the most important contents they can explain a lot of these problems. Hanno van Keulen emphasised that they should know more about practical things, e. g. about pollution and therefore, e. g. how to clean the toilet and what to do with the washing etc. which is also chemistry. He questioned the use of learning about Niels Bohr for people who are e. g. trained to be a butcher. In Florida, for example, the size and the distance of the sun had been measured. One student had been asked why he had done that, and he answered that he did not know. Hanno van Keulen also criticised the example of purifying ester by destillation — the ester will be thrown away afterwards anyway.

The students in America do not like science. It has no meaning for them to know e. g. the temperature of a star, because nobody ever asks them about this. Science is not applied in everyday life. It is an abstract and it only becomes concrete in the laboratory.

In other countries it is not possible to teach new subjects because of the national curriculum. In the Netherlands there is also a tendency to introduce a national curriculum. Until now Hanno van Keulen can change the subjects if he wants to. He described his empirical research method as follows:

At first he thought about students' misconceptions e. g. about reactions. They all thought that it is always A + B → C. This reaction conception hindered them to understand some observations. The educational outcome is that he changed the type of experiments and therefore the observations. Students tend to ignore things that do not fit in their conceptions.

Another example for their misconceptions was the Brønsted acid–base concept which has to be used to explain esterification. The students cannot understand it if they have Lewis' acid–base concept in mind which involves electronegativity.

Teachers have to have a high tolerance of this teaching method. It used to be popular, but it was given up, because too many problems occured. Hanno van Keulen claimed that 'teacher proof experiments' should be developed. The teacher has to learn about chemistry. He should be able to let the students have experiments and free discussions. Therefore, clues should be given to them in the teacher training.

Today there is a change in the theory of didactical research and this might cause a change of the teachers' methods.

It is not Hanno van Keulen's aim to prove that there are certain misconceptions because there is not enough time. He only wants to obtain an idea about what is going on and to find some qualitative arguments to convince the audience, e. g. to consider the reasoning steps from Brønsted to Lewis.

Hanno van Keulen is also interested in the tests of other researchers. For example, an

ERIC

Australian test had been conducted in America and it gained the same result, which means that students are basically the same in other countries. They differ because of their prior knowledge. It is the most important thing to Hanno van Keulen that his students do organic synthesis. Therefore he does not focus on other researchers' conceptual framework. Anyway, it has to be doubted whether it is helpful. There is nothing wrong with misconceptions, because everybody has them, and therefore teachers have to be aware of this fact.

ERIC

STUDENTS' ABILITIES IN WORKING WITH CHEMISTRY TEXTBOOKS

Gerhard Meyendorf, Pädagogische Hochschule Halle–Köthen, Germany

1. Introduction

For more than 25 years our research group in Köthen has been investigating chemistry schoolbooks. The terms "schoolbook" and "textbook" have different meanings. Schoolbooks refer to all books or booklets printed for and used during education at school. Textbooks are one form of schoolbook besides collection of tasks, compendium, etc. The studies have two main subjects: the chemistry schoolbook itself and the students' work with the textbooks in chemistry lessons.

1.1 The chemistry schoolbook itself

Here the research focused on three questions:

- ❑ What are the characteristics of the different books, what are the distinctive features?
- ❑ What relationships exist between typical contents of chemistry lessons and their presentation in textbooks?

Typical contents with specific characteristics are, for example, several elements and compounds, general laws and relations, technical applications of chemistry, historical aspects of chemistry, etc. The teaching of these contents requires the use of different methodical strategies which affect the presentation of such content in textbooks.

- ❑ What effects do the books as a whole and their different elements like text, illustrations, tables, registers, etc., have on the students?

Which forms of presentation are the most comprehensive and useful ones for the students? From our investigations we derive suggestions for the development of schoolbooks.

Questions concerning the selection of particular topics for the textbooks of the different type of schools have not been considered yet. This problem was not relevant in the former GDR because of the standardized curricula that dictated the content to the last detail. Therefore, our studies focused on how the textbooks can be used in practice most effectively.

Methods of these investigations are: analyses and comparisons of books; interviews with teachers and students; single–student and group tests on the effects of different forms of presentation and testing of new forms of presentation.

1.2 Students' work with textbooks in chemistry lessons

Here the research concentrated on the following questions:

❑ What part do schoolbooks play in chemistry lessons? How can they be used in practice?

To avoid misunderstandings it has to be pointed out that we regard a textbook as an instrument for the lessons. The students should not learn chemistry only from textbooks. Experiments will always represent the most important part in chemistry lessons. Textbooks can be used to support the experiments. Pictures, diagrams, tables and summaries can be used as working material in the lessons. With textbooks students can repeat and rehearse certain topics and confirm their knowledge.

Additionally, from our point of view, science textbooks are models which students can use to learn how to work with science literature. This is an important part of chemistry teaching which can contribute to the students' overall education. Our work aims at developing efficient methods for using textbooks in science classrooms.

❑ What abilities in working with textbooks do students have? What are the reasons for their problems with science literature?

The experiences from school investigations suggest that many students are not able to use their textbooks efficiently. This is not necessarily caused by previous education in other subjects. Chemistry textbooks have certain features with which the students have to become acquainted.

❑ How can students' skills in working with literature, and with certain elements of textbooks, be improved?

In this context it also was investigated whether students from higher grades are able to solve complex problems with the help of their textbooks. The results of these studies were to support recommendations on more efficient ways of using textbooks in chemistry lessons.

The methods of the investigations were the monitoring of lessons, tests with students and empirical investigations with large populations. Also, small groups of students were monitored while they were solving their tasks. The problem–solving processes and the test results were recorded.

There are some problems with investigating students' abilities in working with science literature. For example, the experimental conditions that influence the effect of a certain treatment are difficult to control. It is nearly impossible to take all factors into consideration. It is therefore not always possible to conduct empirical investigations in normal school lessons. The conditions that are necessary for the experiment can only be achieved in an artificial environment. Another problem is that the development of certain abilities is an individual process. The tests have to be set in a way that allows the researcher to observe individual students.

The following text only presents a very small part of our investigations that deal with students' abilities in working with science literature.

ERIC

2. Aims of the investigation

The study was to provide some knowledge on how students gather information about a subject they had not dealt with before. This is a feat students will have to perform again and again, and not only in chemistry. The research questions were:

- To what extent can grade 9 students gather information from different literature about a topic they had not dealt with before?
  - How do they find information in literature?
  - How do they process the information?
- How can problems in working with literature be minimized by the teachers?
  - What treatments can be used?
  - What are the effects of these treatments?

The tasks that were assigned to the students were chosen from topics that are difficult to approach by practical (experimental) work.

3. Design

The investigation was conducted in the 9th form which is one year before a large number of students leave school.

Fifteen classes participated in an initial test that took place at the beginning of the schoolyear. Three questions on new topics were to be solved with the help of three schoolbooks ('Lehrbuch Chemie Klasse 9', 'Wissenspeicher Chemie' and 'Wissenspeicher Physik'). Figure 1 on next page shows the design of the study.

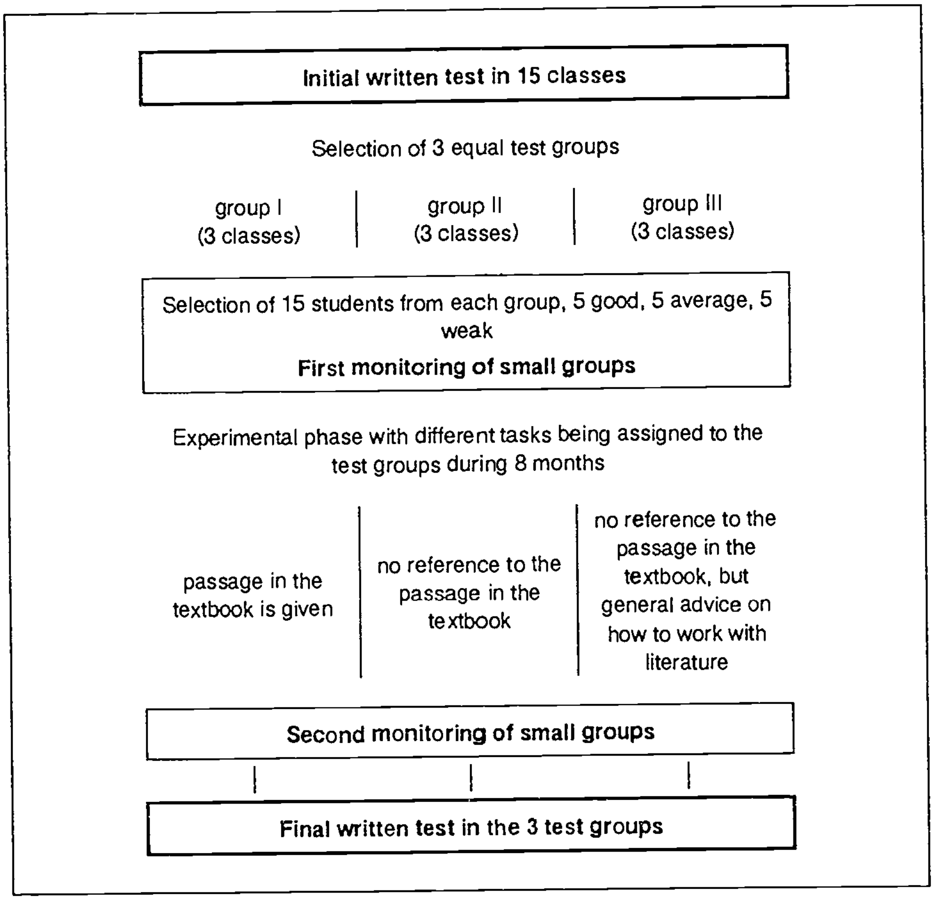
According to the results of the initial test three equal groups each of three classes were formed. Thus, of the fifteen classes, nine eventually took part in the investigation. The three test groups were composed according to the students' grades in the subjects chemistry, biology, physics and mathematics on the one hand and on the results of the initial test on the other hand. The homogeneousness of the groups was validated by using the χ^2–value with regard to the *Brand–Snedecor*–formula (Claus–Ebner, 1967, p. 232). During the schoolyear the students were assigned 7 tasks on topics that had not been dealt with before. The examinees were to solve these tasks with the help of information from their textbooks.

In group I the students were told which passages of which books they could use for solving the task. Thus, these students only had to interpret the information given.

In group II the students were not given such hints. They had to find the passages and the necessary information themselves and to interpret them.

In group III the students were not given hints, either. However, the teachers were asked to give some general advice on how to work with literature. They were asked to refer to the following aspects: analysis of the task; planning a solving strategy; selection of literature; obtaining the information needed from literature; obtaining information from different structural elements; condensing the information to an appropriate answer and checking the answer.

Figure 1: The design of the study



In all test groups, the answers the students noted were evaluated in the intermediate test. At the end of the term, in all test groups the students were asked to work on three tasks which comprised the final test which was similar to the initial test.

The conditions of the experiment were as follows:

- The contents of the lessons were in strong connection to the detailed syllabus. The pedagogical concepts as well as the test items, the topics of the test items and the evaluation system of the questions were based upon the "Unterrichtshilfen". Therefore, these factors were regarded as constant.
- The advice the students received during the test period was considered as an independent variable.
- The quality of the students' results during the tests and observations were regarded as a dependent variable.

From each of the three test groups fifteen students were selected and monitored in small groups. In each case five of the fifteen students had a high, five had an average, and five a low standard of performance. Shortly after the initial test these students were to prepare a paper, the topic for which was set in the lesson "Explain your class mates in a short report how ethanol, also called ethyl alcohol, can be synthesized through fermentation".

The students selected were monitored while they were preparing the paper in groups of two. For this task the following nine different books were at the students' disposal:

(1) Lehrbuch Chemie, Klasse 9. Verlag Volk und Wissen Berlin

(2) Chemie in Übersichten — Wissenspeicher für Klasse 9 und 10. Verlag Volk und Wissen Berlin

(3) Sommer: Wissenspeicher Chemie. Volk und Wissen Verlag Berlin

(4) Ludwig: Allgemeine, anorganische und organische Chemie (Wissenspeicher). Deutscher Verlag für Grundstoffindustrie Leipzig

(5) Lehrbuch der Chemie. Deutscher Verlag für Grundstoffindustrie Leipzig

(6) Grosse / Weißmantel: Chemie selbst erlebt. Urania Verlag Leipzig, Jena, Berlin

(7) Studienmaterial Chemie für die Erwachsenenbildung.

(8) Döhring / Golisch: Grundlagen der organischen Chemie. Deutscher Verlag für Grundstoffindustrie Leipzig

(9) Kleine Enzyklopädie Leben. Verlag Bibliografisches Institut Leipzig

The students worked without assistance. Their attention was drawn to the books and it was made sure that they comprehended the task since it was the first time for them to examine a new subject. The activities of the students were recorded in a protocol which included:

- Selection of the literature used
- Sequence of the literature consulted
- Obtaining the information needed from literature (by skimming through the books, by table of contents, by index)
- Success in searching the information
- Quality of the search (three–grades–scale)
- Choice of catchwords
- Time needed for using the books
- Use of the structural elements in the books
- Profoundness of the use of the literature (three–grades–scale)
- Quality of the information obtained (essential (+), important and unimportant (+ –), only unimportant (–))
- Timing and style of taking notes
- Quality of results (text or notes, structured or not, reference to the subject, correctness)

Single students reported on their results in the following lesson. The students' activities were therefore part of their usual class activities.

Just before the end of the investigation the same small group of students was monitored once more. This time the groups had to report on the subject "Give a general view on the different plastics that are produced by polymerization". The results were evaluated in the same way as in the initial test.

4. Results and discussion

4.1 Comparison of the test groups

Only a few results and interpretations will be described here in order to focus on the monitoring of the small groups. In the initial control the students had to obtain information about three subjects that had not been dealt with beforehand from the textbooks. The subjects were:

- Difference between the term *velocity* in physics and in chemistry
- Le Chatellier's principle
- The term 'reversible reaction'

The pieces of information the students could be expected to find were listed and then tested on other students in a pretest. The percentage of correct answers was determined.

One hundred and ninety nine students from 9 classes gathered 58,9 % of the maximum amount of information. The results of the three groups were as follows:

Table 1: Initial control. Percentage of correct answers

Group	Number of students	Result
I	59	56,3 %
II	69	61,0 %
III	71	59,6 %
Total	199	58,9 %

The observations during the control suggested the following tendencies:

- The majority of the students was able to find the appropriate passages
- quite frequently they did not use the index (which is the most effective way)
- The students often worked superficially with the literature.

From the answers it could be concluded which schoolbook had probably been used. The precidence of the 'Wissenspeicher' (students' lexicon of chemistry) is remarkable, also in combination with the textbook. The textbook itself did not play an important part.

Table 2: Initial control. mainly used schoolbooks

schoolbook used	how often	percent
Textbook	17 times	8,5 %
'Wissenspeicher'	100 times	50,2 %
Textbook + 'Wissenspeicher'	82 times	41,2 %
total	199 times	100 %

In the final test the three groups had to solve the same problems. Here, the three tasks

ERIC

referred to:

(1) the comparison between polymerization and polycondensation,
(2) the primary materials of the synthesis of phenoplastics,
(3) the production of synthetic rubber.

The results in comparison with the pretest are as follows:

Table 3: Results of the final control, compared with initial control

group	initial control	final control
I	56,3 %	67,5 %
II	61,0 %	70,2 %
III	59,6 %	81,7 %

In the final test the groups I and II were approximately 10 % more successful than in the initial test. This can be due to the practice during the term. Another reason might be that the tasks were less difficult. The result of group III was much better with an improvement of over 20 %.

There was no difference between the results of groups I and II. Finding the adequate passages did not seem to be a big problem for the students. However, students who had been trained to find the adequate passages were more competent in working with literature afterwards. The members of group III were much better at obtaining information. Observations during the test showed that these students were more effective and more confident in working with literature. They worked more thoroughly.

Two interesting results can be derived from this fairly rough comparison:

- Students often used the textbooks ineffectively and superficially. Finding the information was not very difficult for the ninth form students whilst quite often not all of the essential details were drawn from the text. However, this is a basic requirement for successful practising and repeating at home.
- It turned out to be advisable for the teachers to deal with the methods of working with literature in the lessons from time to time. During the investigation (one school year) the teachers did this seven times.

4.2 Results from monitoring small groups

The monitoring of small groups was supposed to provide more detailed information on how students work with literature. The initial observation showed the situation before the treatment. The second observation took place eight months later to provide a general idea about what effects the students' work had.

4.2.1 First observation

The pieces of information the students gained from literature when they were preparing the reports were compared with a list of all passages that could be expected to be found: 301 bits of information that were important for the reports were collected from the literature by the 45 students. This information is 62.8 % of the total amount of important literature passages. In addition 119 'incorrect', that is unimportant, pieces of information were gathered. The percentage of important pieces was approximately the same as in the initial test for the whole group (Table 4). This seems to show that the sample of 45 students was homogeneously drawn from the 9 classes, even though the tasks were on different topics. The distribution among the groups I, II and III appeared to be homogeneous.

Table 4: Percentage of important bits of information, first observation (N = 45)

group	pretest (N = 199)	total	by quality of students		
			good	average	weak
I	56.3 %	57.8 %			
II	61.0 %	65.0 %	73.9 %	66.3 %	48.1 %
III	59.6 %	65.0 %			

Table 5 shows what characteristics the unimportant ('incorrect') pieces of information had. The high percentage of double statements and answers not related to the task is remarkable. This suggests that many students did not compare their results with what was requested in the task.

Table 5: Number of 'incorrect' pieces of information gained during the monitoring of small groups by quality of students' performance

type of 'incorrect' information	total	good	average	weak
double statements	34	12	8	14
unclear, misunderstood	14	7	3	4
not related to the task	46	18	12	16
incorrect, incomplete	25	9	8	8
sum	119	46	31	42

Furthermore, it appeared that there were no essential differences concerning incorrect statements between the three groups. This is in accordance with the results of other investigations. The students' abilities in examining chemical contents often do not correlate with their abilities in working with and informing themselves from teaching aids such as textbooks. (Also, it seems that their abilities in this area are not being rated which is also true for experimental work). Perhaps the average students, who gathered the least amount of unimportant (incorrect) information, use their textbooks at home more frequently and are therefore better at working with them than the students with good marks who perhaps do not have to use textbooks very often.

Half the notes students produced during their literature work on the reports were written as text, the other half consisted of catchwords. Only a few students structured the information well and arranged them in a sensible sequence.

While nine books were available for the students, most students, however, used 3 to 5 books. In 75 % of the cases the students used, as expected, the 3 books they knew from the lessons. Students success in working with the books can be seen from Table 6. The number of unsuccessful attempts is remarkably high although all books did contain information that was useful for the reports. What are the reasons for these unsuccessful attempts? A first reason is that many students had problems with the methods of searching for information. A search for information was altogether performed 262 times.

Table 6: Students' success in working with the books

Total use of literature	248 times
a) singular use of one book	211 times
b) repeated use of books	37 times
Unsuccessfully used (no information on the subject was obtained)	102 times
Successfully used	146 times
a) to get new information	100 times
b) to check results	46 times

Searching via the index was chosen quite often (155 times, i. e. 59.2 %) which can be regarded as positive. However, Table 7 also shows that in 188 cases (71.6 %) the quality of performance was assessed as average or less weak.

Table 7: Students' methods of collecting information

searching method	total	quality of performance		
		good	average	weak
via table of contents	71	2	29	40
via index	155	72	68	15
via skimming	36	–	3	33
total	262	74	100	88
			188	

It can be seen from Table 8 that the elements were not thoroughly used. This also leads to the result that the optimal amount of information is not obtained. If the summaries which contain the essential information are not considered, the difficulties of the ninth form students in working with literature becomes apparent. Only in 16 % of the cases did the students manage to gain important information from the texts.

Table 8: Use of literature elements

element	total	thoroughly used	super-ficially used	Information gained		
				important	important and unimportant	only unimportant
table	5	1	4	2	–	3
diagram	6	1	5	–	3	3
summary	57	5	52	33	20	4
text	50	10	40	8	22	20
total	118	17	101	43	45	30

ERIC Full Text Provided by ERIC

The results of the first observation are:

- The ninth form students were able to gain single bits of information from the books, but they had problems with choosing the essential bits.
- The interpretation of text(s) was particularly difficult for them.
- Finding the information was very difficult. The students chose ineffective ways or they did not use the possibilities offered in the books.
- They also had problems with arranging the information according to the task.
- Some difficulties were caused by misunderstanding the task.

4.2.2 Second observation

The second observation was to test the effects of the students' experience with literature that was gained throughout the school year. Also, it was to confirm certain tendencies that were found during the first observation. The students were not influenced very much during the 8 months so that only small improvements could be expected.

The results of both observations are compared in Tables 9 – 13. Table 9 shows 'correct' and 'incorrect' pieces of information.

Table 9: 'Correct' and 'incorrect' pieces of information

group	'correct' bits of information (% of all possible bits)		'incorrect' bits of information (number of bits)		difference
	1st observation	2nd observation	1st observation	2nd observation	
I	57.8	63.2	40	33	– 7
II	65.0	65.4	42	35	– 7
III	65.6	73.2	37	14	– 23

There were slight improvements in all test groups. The improvement can particularly be seen in group III which had received instruction on how to use literature efficiently. The effect can also be seen from the decrease in the 'incorrect' bits of information. In all, the 45 students used the books 354 times. As shown in Table 10, the proportion of successful attempts increased, especially in groups II and III. These groups had to gain information on their own during the investigation.

Table 10: Successful and unsuccessful attempts at finding information in the books, percentage of attempts

group	successful		unsuccessful	
	1st observ.	2nd observ.	1st observ.	2nd observ.
I	57.7 %	71.7 %	42.3 %	28.3 %
II	57.0 %	95.7 %	43.0 %	4.3 %
III	61.2 %	95.3 %	38.2 %	4.7 %

In the searching methods there was a stronger tendency towards searching via the in-

ERIC

dex. The proportion of less efficient methods decreased as shown in Table 11. As presented in Table 12, the quality of the students' performance, especially in groups II and III increased, with the increase being approximately 30 %.

Table 11: Students' searching methods

group	via index		via table of contents		via skimming	
	1st obs.	2nd obs.	1st obs.	2nd obs.	1st obs.	2nd obs.
I	46.0 %	63.1 %	31.6 %	21.0 %	22.4 %	15.8 %
II	63.9 %	68.0 %	24.4 %	21.3 %	11.6 %	10.6 %
III	65.0 %	74.6 %	26.0 %	16.4 %	9.0 %	8.9 %

Table 12: Quality of students' performance

group	good		average		weak	
	1st obs.	2nd obs.	1st obs.	2nd obs.	1st obs.	2nd obs.
I	25.0 %	36.8 %	34.2 %	38.6 %	40.8 %	24.6 %
II	26.8 %	57.4 %	41.9 %	29.9 %	31.4 %	12.7 %
III	32.0 %	62.7 %	38.0 %	32.8 %	30.0 %	4.5 %

Once more it was estimated whether the students worked thoroughly or superficially. Also, it was estimated whether essential, important and unimportant or only unimportant information had been gained. It is not possible to compare particular elements of the textbooks because, for example, one book contains more tables but less diagrams than the others. However, the comparison of the data suggests that the students' thoroughness and their ability to identify the essentials improved.

Table 13: Percentages of students' work with elements of the books (estimation)

students who	1st obs.	2nd obs.
worked thoroughly	14.4 %	33.6 %
worked superficially	85.6 %	66.4 %
grasped essential information	36.4 %	43.0 %
grasped essential and unimportant information	38.1 %	46.7 %
grasped only unimportant information	25.4 %	10.3 %

The second observation showed that:

- instructions for the effective work with textbooks that are given in the lessons from time to time have a positive effect on the result.
- unaided work with textbooks occasionally leads to positive effects as well.
- especially the students' abilities in finding information could be improved by asking them to work with textbooks. This aspect is very rarely considered in the lessons.

ERIC

- during the investigation the students became more critical in selecting information from the books. Furthermore, their ability to select the essential elements with regard to what is required in the task improved.

5. Conclusion

The results of many other studies on students' abilities in working with chemistry textbooks are in agreement with the present study.

The majority of students is able to obtain information from the textbooks in chemistry lessons. However, it is often done superficially and not systematically, and it also leads to incomplete results. This can be avoided by giving instructions on how to work systematically.

Almost all of the similar investigations showed that students have difficulties in solving problems using textbooks. This is not only due to the difficulties with obtaining information, but it is also due to inadequate abilities in processing information. The essential information can only be gained by comparing, arranging, condensing, etc., but many students do not know how to use these methods. Often they do not obtain information because they cannot find an appropriate catchword which is due to a lack of basic knowledge in chemistry.

The students are able to work with pictorial descriptions whereas they have difficulties in dealing with diagrams. This is a method of representation that often is very important for their later profession. Another problem is dealing with tables. The reason for this seems to be that they are not familiar with the basic characteristics of tables (i.e. combination of simple criteria, relationship between lines and columns).

Many teachers suppose that the students are able to deal with the pertinent literature. They assume that students manage to use the textbooks for their homework in a most effective way. This presumption leads to a 'vicious circle':

- Students have problems with the subjects of the lessons;
- They are supposed to close the gap at home;
- They are not able to do their homework properly;
- Their difficulties increase.

6. Discussion of the method

The present empirical investigation revealed some interesting results. However, one has to be aware of certain limitations:

(1) The test population was rather small. Therefore, the results should not be generalized. The present study is to be seen as a pilot study for a major investigation.

ERIC

(2) Some limitations result from the fact that subjective estimations (e. g., quality of performances) were considered in addition to exact data (e. g., pieces of information that were found/not found). One person was responsible for estimation in all the cases. This might have increased the subjectiveness since it was not possible to consider different opinions.

(3) The significance of differences were statistically validated using the χ^2–test according to the *Brand–Snedecor* formula and to the *Kolmogorow–Smirnov* test. These tests were not applied to the subjective estimations.

7. Acknowledgements

I would like to thank Miss Kerstin Erxleben for her support. Unfortunately, she died from an incurable disease after the investigation had been finished. Also, I have to thank to Mr Treagust and a student who supported the translation of this paper.

References

Claus, G. & Ebner, H. (1967). *Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen*. Berlin.

Erxleben, K. (1984). *Untersuchungen zur Befähigung der Schüler zum komplexen Arbeiten mit der Literatur (im Chemieunterricht)*. Dissertation A. Köthen: Päd. Hochschule.

Meyendorf, G.. *Zum Einsatz des Lehrbuches im naturwissenschaftlichen Unterricht*. Wiss. H. d. PH Köthen, vol. 7 (15), H. 2, p. 13ff.

Meyendorf, G. (1986). *Untersuchungen zur Arbeit mit Schulbüchern im Chemieunterricht*. Inform. z. Schulbuchfragen. Berlin. vol. 54, p. 32ff. (Summary with references to other articles)

Schulbuch und Schülertätigkeit — Materialien der IX. Wiss. Konferenz der Forschungsgruppe Schulbuch an der PH Köthen im November 1984 (1985). Inform. z. Schulbuchfragen Berlin, vol. 51, expecially p. 9ff, 29ff, 131ff, 136ff

COMMENTARY OF THE DISCUSSANT

Heinz Schmidkunz, Universität Dortmund, Germany

Although I have not investigated the effect of pupils textbooks on the learning of chemistry I had to realise all the problems by writing such a book. Many years ago I wrote a chemistry–textbook together with another colleague for the German "Realschule", that means for pupils at the age of 13 to 16 years. At this age usually the education in chemistry is started.

The main question by writing such a book is to find the purpose the book should be used for. There are several possibilities and it is not easy to combine all these aspects in one textbook.

The possible forms of application are the following:

- ❑ Should it be a textbook to give complete information concerning the subject chemistry?
- ❑ Should it be a book supplementing the chemistry lessons at school, a book for the use beside the lessons (for example at home)? For example to fill up blanks in knowledge?
- ❑ Should it be a book for using throughout the chemistry lessons at school, characterised as an exercise-book?
- ❑ Should it be a book for repeating chemistry which was taught at school?

As you may imagine it is very difficult to unify all these aspects into one book. There are more principal questions concerning the layout and the contents of the books.

- ❑ The relationship between typical contents of the chemistry-teaching and its representation in the textbook is of great importance. That concerns every-day-chemistry, environmental chemistry, significant processes in chemical industries, the meaning of history for the teaching process and not unfamiliar topics of pure chemistry.
- ❑ Can the topics be used in later situations?
- ❑ Is the chemical content intelligible enough for the pupils or students?
- ❑ Do the demonstrated experiments (in pictures or described) work? Are they easy enough and practicable? Is there a significant relationship between the experiment and the theoretical background?
- ❑ Which concept is used by developing chemical terms? Is a distinct method of chemistry-teaching applicated in the textbook? For example is there found a method of problem-gaining and problem-solving derived from pupils' interests?
- ❑ The scientific facts must be represented in an understandable way (a problem of didactic reduction).
- ❑ The text should be written in a readable manner, clear expressions and short phrases are desirable.
- ❑ The relationship between the text-part and the picture-part should be well balanced.
- ❑ The text should be an assistance for the pupils to construct their own knowledge.
- ❑ For a powerful stimulating learning the laws of perception should be realised in the book. That includes the pictures as well as the written part.
- ❑ To summarise: textbooks are better if their layout is done in a clear cut fashion adapted to the age of the learners.

In this context another point of view is of great interest: Using literature for students is necessary and a kind of learning on their own. So we have to ask and we are wondering if they are able to do so. The investigation of Gerhard Meyendorf has confirmed our experience that students are to be carefully led to literature, step by step. So working with literature within chemistry lessons and generally in chemical education should be scheduled by the teacher or already be fixed in the curricula. As a further result of the investigations of Gerhard Meyendorf may be concluded that students will comprehend the utility of using literature in learning chemistry and solving problems. So using literature is a part of chemical education. That will be the reason for providing suitable literature and time enough for exercises in using.

With regard to the literature itself some alterations have to take place. As it was shown in the investigation pupils get difficulties in quickly finding the right terms or processes they are looking for.

A clear and well readable register is granting a good survey. In addition to that an extended keyword–index assists to find the right terms, words, compounds or anything else quickly. In any way a good orientation is necessary, meaning that even the pages are to be marked more precisely as it is common in use. Many pupils and students are visual types, the layout of the pages remains on their minds and when reaching for a book, they at first prefer looking at the pages and not at the register.

Summarising it may be mentioned that students must gradually be led to use literature. At first to the textbook which is provided for the lessons and later on to other kinds of books concerning the subject. That may occur within the lessons at school or at home.

On the other hand the authors and editors of textbooks have to provide a good and quick orientation. Register and keyword–index should carefully be developed and the pages must get a characteristic shape and layout in relationship to the subject–matter.

SUMMARY OF THE PLENARY DISCUSSION

Petra Beuker, Universität Dortmund, Germany

According to the contribution of the discussant many problems occur when writing a chemistry textbook. It is impossible to write a textbook that all teachers like, because different styles of teaching require different books. One single book cannot be suitable for all situations.

A lot of time is spent on what to learn and not on how to learn which is more important. One of the participants who has written a textbook himself said that he has not thought deeply about how students use it.

It depends on the books whether the table of contents or the index is used. If the students find information by using the table of contents they will not use the index. The first strategy which is successful will be kept.

It is easier for them to look up 'plastics' in the table of contents, for example. Small pieces of information and details are given in the index. The reason for not getting it on the index is not necessarily the poor quality of the book. The use of catchwords is only helpful if the students have some knowledge about the system of chemistry. If the teacher wants the students to search for some information in the book he usually tells them the pages etc.

There has been an investigation about students' abilities in working with textbooks at the end of 10th class. Two years later the same students who have now changed to a professional school had been investigated in the same way again. The result was that there had been no increase in their abilities.

In the list of literature there is only one official textbook. The different groups in Gerhard Meyendorf's investigation all used the same books. During the investigation the students came in groups of two to the investigator. The nine books were available for them. They knew some of the books from school. The students did not work in teams.

As for the coding of the data–collection, the initial and the final test were written in the class like an exercise. In the pre–test the answers were compared to expected answers. It was only counted how many times which answer occurred. The monitoring in the intermediate control was done as follows: A protocol was prepared that included six points which were marked with symbols. For example, how the students begin or how they grasp the task. The investigator had to make sure that they understood the task. The process of searching for information was categorised into a) successful, b) not successful or c) successful and not successful method of gaining information. It had also been monitored whether they were skimming or using the table of contents. The quality of their searching method was estimated as 'good', 'average' or 'weak'. Another point of interest was whether they wrote down information and whether they solved the task by using the elements of the textbooks.

The aspect of subjectiveness of the coding was also discussed. Not more than two students could be monitored at the same time. The investigator did not ask questions because it would have disturbed the students and probably they would have got useful information from that. Therefore they were only monitored. It had been reconstructed from their answers which textbooks they might have used.

There was only one researcher monitoring. One advantage of this is that there are no different opinions and impressions. A disadvantage is the subjectiveness.

Researching on the development of ability in working with chemistry textbooks is difficult because it is more than only investigating chemical contents. It is very difficult to prove Gerhard Meyendorf's results with statistical methods.

META–ANALYTIC AND MULTIVARIATE PROCEDURES FOR THE STUDY OF ATTITUDE AND ACHIEVEMENT IN SCIENCE

Michael D. Piburn, Arizona State University, USA

Abstract

This research involves the use of meta–analysis to identify factors that are predictive of success in science. Results from many studies are reduced to a single variable, the correlation coefficient, thus allowing the use of multivariate statistics.

Dependent variables that offer operational definitions of success in science include performance in a variety of subject areas and grade levels. Independent variables are of three types. The first consists of a group of ability and neo–Piagetian variables that are primarily psychological. A second concerns background and preparation in the sciences, and is generally characterized in the literature under the rubric of background knowledge. The final group is those societal indicators that are grouped to form the concept of socio–economic status.

Data collected for this study suggest that the relationships of socio–economic status and attitude to achievement in science are weak. Among other variables, the most important precursors of scientific achievement appear to be spatial ability, memory capacity, and background knowledge. These results are fully consistent with those in the literature of expertise.

1. Introduction

The procedure of meta–analysis was suggested by Glass (1976) as an alternative to other methods then in use for the review of prior research. It a powerful means of aggregating quantitative results across a large number of research reports, and has commonly been used in summarizing the results of experimental studies. In the most often employed technique, effect sizes (the difference between the means of control and experimental groups divided by the pooled variance) are computed for each study, and averaged for the purpose of the review. Less commonly, meta–analysis has been used to compare the results of correlational studies. While a variety of procedures are available for weighting the values of correlation coefficients from different studies (Hedges & Olkin, 1985; Schmid, Koch, & LaVange, 1991), these have not been used in the few studies of this type to be found in the science education literature. Instead, the procedure of choice has been to collect a pool of similar correlation coefficients and to report their means and variances.

This study was undertaken to test the feasibility of using meta–analysis to complete a correlation matrix from previously reported research that could then be used to conduct a secondary analysis. Specifically, it was the purpose of this study to complete a multivariate statistical analysis of factors that are predictive of success in science.

Dependent variables that offer operational definitions of success in science tend to vary by subject area, grade level, and type of examination. Thus, for example, one might read research that reports performance on gas law problems by college freshmen, laboratory applications by high school students, or science process items by elementary school students.

Independent variables are also of several types. An important group contains primary psychological constructs, of which a very large number have been examined. These might include, for example, IQ, digit span, or spatial visualization. A second set of independent variables concerns background and preparation in the sciences. Included are previous course work and grade point average, prior skills and knowledge, and misconceptions. These are generally characterized in the literature under the rubric of background knowledge.

A third group contains all social indicators, such as socio–economic status, affiliation with a minority group, parental education, or home environment, that are known to be connected with school success. It is not unreasonable to expect that these will work together in the same general way in their capacity to predict success in science.

To summarize, independent variables that are being considered in this study cluster approximately into three groups; psychological, experiential, and societal. Put in more familiar terms, these are roughly characterized as ability, background knowledge, and social class. Dependent variables are traditional achievement measures, reflecting either school success as measured by examination score and grade achieved, or by standardized local oı national tests of scientific achievement.

2. Methodology

In order to conduct multivariate statistical procedures it is necessary to have a complete matrix of correlation coefficients between all variables to be considered. If any portion of that matrix is absent, the variables(s) in question must be eliminated from the analysis. All issues of the *Journal of Research in Science Teaching* from 1983 through 1992 were reviewed, and a total of 51 articles were found which contained useful data. In most cases, correlation coefficients were extracted directly from the article. However, it was occasionally necessary to record a regression coefficient instead. Such coefficients "can be interpreted much like an ordinary coefficient of correlation" (Kerlinger, 1973. pg. 621). In a few studies an unusually large number of similar correlations were recorded, as for example the relationship between a variable and 3 – 5 separate examination scores in several different courses. In such cases, where it seemed suitable, a single average was computed and recorded.

The technique used in this study is regression analysis. Two procedures for regression are available, and both are used. In the first, hypothesis testing is conducted by varying the order of entry of independent variables into the regression equation. This follows the injunction by Kerlinger that, since order of entry has a profound impact on increase in variance explained at each step, "order of entry of independent variables into the regression is determined by the research problem and the design of the research" (1973, pg. 628).

If there is no theory-based reason for ordering the entry of independent variables, interpretation is more easily based upon the values of Beta. These are the standardized regression coefficients, whose magnitude is independent of the order of entry. They can be thought of as equivalent to simple correlations between two variables (Nie, Hull, Jenkins, Steinbrenner & Bent, 1975, pg. 325). Under normal circumstances, Beta coefficients are subject to the same type of significance testing as correlation coefficients. However, this does not appear to be so easy in a meta-analysis, where the correlation coefficients used in the regression are not associated with any sample size.

The analyses reported here were conducted on an Apple IIGS computer with the use of a program titled *Statistics With Finesse* (Bolding, 1985).

3. Results

Publication over a ten-year period of the *Journal of Research in Science Teaching* yielded 51 articles which contained a total of 262 usable correlation coefficients. These were grouped into 70 different categories, and summary statistics were computed for each.

From among these, 14 represented relationships between achievement in science and other variables (Table 1). They appear to constitute the psychological, experiential and societal factors that this study was designed to examine. Achievement measures included test and examination grades, gain scores from pre- to post-test, course grades, grade point average, and achievement on standardized tests.

Table 1: Mean correlations between scientific achievement and background variables

Achievement in science	Mean correlation	Standard deviation	Number of studies
General ability	0.41	0.22	5
Verbal ability	0.40	0.25	4
Spatial ability	0.41	0.22	5
Cognitive level	0.44	0.15	15
Critical thinking	0.60	0.29	3
FDI	0.29	0.18	13
Locus of control	0.21	0.21	2
Mental capacity	0.21	0.20	2
Prior knowledge	0.39	0.12	8
Quantitative reasoning	0.35	0.22	14
Scientific reasoning	0.40	0.16	6
Attitude	0.31	0.14	14
Math Anxiety	−0.21	–	1
SES	0.33	0.17	4

Psychological factors reported in the literature were quite varied. However, in order to conduct the statistics reported in this study it was a requirement that a complete correlation matrix of all measures with one–another be compiled, and this was not possible in all cases. For example, Rotter's Locus of Control and the Watson–Glaser Critical Thinking Appraisal had to be dropped, despite their obvious interest. In the final analysis, 6 psychological variables were retained. These were general ability, verbal and spatial reasoning, field dependence–independence, mental capacity, and cognitive developmental level. A factor analysis of these was conducted, and they all clustered into a single factor with loadings of between 0.53 and 0.81, and an eigenvalue of 3.19. No other factor with an eigenvalue of more than 1.00, the normal default option, could be computed. From this it was concluded that all of the psychological variables came from a single psychometric pool of items, and that there was no statistical basis for their classification. Thus they were organized into two groups on the basis of a priori theoretical constructs; ability and neo–Piagetian.

Three general categories of background or prior knowledge were aggregated. Two were measures of scientific and quantitative reasoning skills, and the third was constructed entirely from variables which, in the original study, had been characterized as measuring prior knowledge. This last group ranged widely, including pre–tests, standardized achievement tests, prior course work, and prior Grade Point Average.

All attitude measures were summed into a single pool, as were all measures of socio–economic status.

3.1 Comparison with prior meta–analyses

Three meta–analyses of the relationship between achievement and other factors were completed and reported in 1983. The work of Steinkamp and Maehr (1983) was concerned primarily with gender differences, and their analyses were conducted separately for males and females. For example, they reported correlation coefficients between cognitive ability and achievement of 0.36 for males and 0.32 for females. This is slightly lower than the value of 0.44 for this same relationship reported here. For achievement and attitude, they reported correlation coefficients 0.19 and 0.18 for males and for females. This is similar to the value of 0.16 obtained by Willson (1983), and both are somewhat lower than the value o. 31 for the same relationship reported in this study.

One study (Fleming & Malone, 1983) reports a greater variety of data, is more comparable to, and shows results that are much more like those reported in this study (Table 2). The differences among these data are greatest in those cases where the number of correlation coefficients reeported is smallest. However, they are in general quite similar, and yield confidence not only in the stability of the relationships through time, but also in the technique of meta–analysis itself.

Table 2a: Comparison of results of meta-analysis by Fleming and Malone and of this study. General ability.

General ability	Fleming & Malone	This study
Cognitive level	r = 0.38 SD = 0.24 n = 11	r = 0.38 SD = 0.20 n = 6
Science achievement	r = 0.43 SD = 0.22 n = 42	r = 0.41 SD = 0.22 n = 5
Science attitudes	r = 0.15 SD = 0.16 n = 13	r = 0.26 SD = 0.01 n = 2

Table 2b: Comparison of results of meta-analysis by Fleming and Malone and of this study. Language ability.

Language ability	Fleming & Malone	This study
Cognitive level	r = 0.30 SD = 0.31 n = 3	r = 0.47 – n = 1
Science achievement	r = 0.41 SD = 0.16 n = 5	r = 0.40 SD = 0.25 n = 4
Science attitudes	–	r = 0.10 SD = 0.20 n = 2

Table 2c: Comparison of results of meta-analysis by Fleming and Malone and of this study. Mathematics ability.

Mathematics ability	Fleming & Malone	This study
Cognitive level	–	r = 0.50 SD = 0.09 n = 4
Science achievement	r = 0.42 SD = 0.19 n = 0.13	r = 0.35 SD = 0.22 n = 14
Science attitudes	–	r = 0.21 SD = 0.19 n = 3

Table 2d: Comparison of results of meta-analysis by Fleming and Malone and of this study. Socio-economic status.

Socio-economic status	Fleming & Malone	This study
Cognitive level	–	–
Science achievement	r = 0.25 SD = 0.09 n = 21	r = 0.33 SD = 0.17 n = 4
Science attitudes	r = 0.03 SD = 0.11 n = 13	r = 0.07 SD = 0.02 n = 2

3.2 The effects of ability

Despite recent movements to subdivide the intelligence concept, such as the Sternberg's Triarchic Model or Gardner's Multiple Intelligences, Spearman's g (general intelligence) is alive and well. Arthur Jensen, one of the concept's more forceful contemporary advocates, believes that this substrate of intelligence shares more variance with a greater range of cognitive activities than any other single factor (Sternberg, 1990).

Five relationships between general ability and achievement in science were obtained in this study. The ability measures used were the abstract reasoning sub-test of the Differential Aptitude Test (DAT), the Primary Mental Abilities Test, Raven Progressive Matrices, the School and College Abilities Test, and the Otis-Lennon Intelligence Test. The mean correlation between achievement and ability for this group was 0.41.

It has been well accepted for almost a century that if any two types of ability are factorially distinct from one-another, they are verbal and spatial (Lohman, 1988). This has been given further credence more recently by the research of Sperry (1961) with commissurotomy patients, demonstrating the very different functions of left and right cerebral hemispheres, and studies of the types of solution to three-term series problems used by visual and analytic problem solvers (Sternberg, 1980). Again, however, g tends to absorb by far the greatest variance in any predictive equation, and the addition of terms for verbal and spatial ability often adds little to its explanatory power.

Spatial ability is especially difficult to define because, although many measures appear to be spatial in character, few of them cluster heavily into a single factor solution. In addition, the demonstrable contribution of spatial factors to achievement is often low. Indeed, Lohman states that "spatial tests add little to the prediction of success in traditional school subjects, even geometry, after general ability has been entered into the regression (1988, pg. 182).

The most commonly accepted primary components of spatial ability are visualization and spatial orientation (Ekstrom, French, Harman & Dermen, 1976). Three of the five relationships found for this study were with spatial rotations. The remaining two were between achievement and the spatial and mechanical reasoning sub-tests of the Differential Aptitude Test (DAT). The mean correlation between spatial ability and achieve-

ERIC

ment was 0.41.

Verbal abilities were measured in four studies, and their correlation with achievement computed. The measures used were the vocabulary sub–test of the Stanford Achievement Test, the verbal sub–test of the Cognitive Abilities Test and the Descriptive Test for Language Skills. Although none are counted among the more traditional measures of verbal ability, they seem suitable for the purpose addressed in this study. The average correlation between these variables and achievement in science was 0.40.

Sufficient information was gathered in the course of this study to conduct a multiple regression analysis of the impact of general, verbal and spatial ability on achievement in science (Table 3). Only one correlation was missing, that between verbal and spatial ability, and a value of 0.34 was obtained from Lohman (1988, pg. 194).

Table 3: Regression of achievement against general, verbal and spatial ability (* see text)

	(1)	(2)	(3)	(4)
(1) General ability	–	0.74	0.53	0.41
(2) Verbal ability	0.74	–	0.34*	0.40
(3) Spatial ability	0.53	0.34*	–	0.41
(4) Achievement	0.41	0.40	0.41	–

Dependent variable = achievement:

Independent Variable	Multiple r	Multiple r–squared	Beta
General ability	0.410	0.168	0.076
Verbal ability	0.434	0.189	0.247
Spatial ability	0.497	0.247	0.286

In a test of previous assertions of the relative importance of these three variables, achievement was regressed against general ability, followed by verbal and then spatial measures (Table 3). This order of entry was specifically chosen to test the claim that the variance in success in school subjects is largely absorbed by measures of general intelligence.

The variance in achievement shared with general ability was 17 %. This was increased to 19 % with the entry of verbal ability and to 25 % with the entry of spatial. The increase in explained variance with the entry of the spatial ability variable was quite remarkable, as were the relatively large values of Beta associated with both verbal and spatial reasoning. From this result it is necessary to conclude that the impact of spatial ability on achievement is much larger than has been previously suggested.

3.3 Neo–Piagetian factors

The publication, in 1958, of *The Growth of Logical Thinking From Childhood to Adolescence* by Barbel Inhelder and Jean Piaget was a significant event for science education. Within a very short period of time this work had caught the attention of science educators, and ultimately led to more than a decade of research within the Piagetian paradigm. The product was a re-consideration of the psychological basis of science education and the acceptance of a constructivist position.

Modern theory in this area is described as neo-Piagetian, and involves an attempt to unify several separate psychological traditions. These include earlier visions such as functionalism and structuralism as well as more contemporary models of information processing and artificial intelligence. As is often the case, this effort has not gone smoothly (Beilin, 1987).

Science educators involved in this new synthesis have been most influenced by the work of Pascual-Leone (1969), which emphasizes particularly the importance of two performance factors, M-demand and field effects, in the completion of Piagetian tasks. If the subjects' mental capacities (M-space) are not adequate to the M-demand of the task, or if they are distracted by field effects, they will not be successful even if they are fully competent in the logical demands of the task.

Mental capacity is most often measured by means of digit span tests, in which a subject is asked to repeat strings of letters or numbers. However, Burtis and Pascual-Leone created a measured called the Figural Intersection Test specifically to measure M-space. In addition, Lawson (1985, pg. 582) has claimed that the Raven Progressive Matrices Test, although most commonly thought of as a pure measure of g, is a measure of mental capacity, and has used it in that fashion (Niaz & Lawson, 1985). Although the concept of field-ground is an old one in psychology, the field effects emphasized in Pascual-Leone's theory refer more specifically to the phenomenon of field-dependence/independence (FDI) formulated by Witkin (Witkin, Dyk, Faterson, Goodenough & Karp, 1962). Witkin's original work, conducted with subjects in an inclined room, characterized people on a continuum from those who were influenced most heavily by internal (the force of gravity) to external (the room, or field) cues. Those latter individuals were called field-dependent. Subsequently, Witkin turned to the Embedded Figures Test to measure this same quality, which he then called restructuring. Those subjects who were unable to restructure were unsuccessful on the Embedded Figures Test and were thus field-dependent.

The Embedded Figures Test is similar to the Hidden Figures Test, which itself is an adaptation of the older Gottschaldt Figures test popularized by Thurstone (Ekstrom, French, Harman & Dermen, 1976). Both of these latter instruments are traditionally considered to be measures of flexibility of closure, which some authors consider to be an element of spatial ability and others contend is related to the ability to break set (Lohman, 1988).

Fortunately, the data set is almost adequate for examining this question. A full matrix of correlation coefficients exists with the exception of that between the Raven and spatial ability. In order to complete the analysis, the correlation of 0.53 between general and spatial reasoning was substituted. Then cognitive level was regressed in two separate analyses, first against spatial ability and FDI (Table 4a, next page) and then against the Raven Progressive Matrices and mental capacity (Table 4b, next page).

In the first instance, spatial ability accounted for 32.5 % of the variance in cognitive level, and this was increased by only 0.3 % with the subsequent entry of FDI (Table 4a). Reversing the procedure by entering FDI first (not shown) yielded a multiple r-squared of 0.16 at the initial step in the equation. The interpretation to be placed on these results is that all of the variance in cognitive level explained by FDI is also contained within spatial measures but that those same spatial measures explain about twice as much of the variance in cognitive level as does FDI.

Table 4: Regressions of cognitive level against (a) spatial ability and field dependence/independence and (b) Raven Progressive Matrices and mental capacity (* see text)

	(1)	(2)	(3)	(4)	(5)
(1) FDI	–	0.47	0.62	0.37	0.40
(2) Raven	0.47	–	0.53*	0.40	0.51
(3) Spatial	0.62	0.53*	–	0.10	0.57
(4) Mental	0.37	0.40	0.10	–	0.37
(5) Cognitive	0.40	0.51	0.57	0.37	–

(a) Dependent variable = cognitive level:

Independent	Variables Multiple r	r–squared	Beta
Spatial	0.570	0.325	0.523
FDI	0.573	0.328	0.076

(b) Dependent variable = cognitive level:

Independent	Variables Multiple r	r–squared	Beta
Raven	0.510	0.260	0.431
Mental	0.541	0.293	0.198

In the second instance, entry of the Raven first into the equation yielded a value for shared variance of 26 %, which was increased by 3 % with subsequent entrance of mental capacity (Table 4b). As in the previous instance, the procedure was reversed (not shown) and mental capacity entered first, yielding a multiple r–squared of 0.14. As before the interpretation is that the Raven shares about twice as much variance with cognitive level as does mental capacity, and virtually all of the variance shared between cognitive level and mental capacity is also contained within the Raven.

The question remains of whether or not cognitive level itself contributes to achievement beyond that explained by the variables already examined. To test this question, achievement was regressed against both neo–Piagetian and ability variables, and then against cognitive level (Table 5 on the next page). In this instance, it was necessary as before to substitute a value of 0.34 for the correlation between general and verbal ability (Lohman, 1988, pg. 194). In addition, it was necessary to assume that the correlation of mental capacity with verbal ability was approximately the same as with general ability.

In earlier analyses, variables were entered in particular order so that specific hypotheses could be tested. This was not true in this analysis, and thus the results (Table 5) are more readily interpreted by examining the values of Beta, the standard partial regression coefficients. The strongest values of Beta are associated with spatial ability and cognitive level, followed by substantially lower figures for general and verbal ability. The Betas associated with FDI and mental capacity are so low as to have virtually no meaning.

Full Text Provided by ERIC

Table 5: Regression of achievement against all psychological variables (* see text)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) General ability	–	0.74	0.53	0.48	0.32	0.38	0.41
(2) Verbal ability	0.74	–	0.34*	0.42	0.32*	0.47	0.40
(3) Spatial ability	0.53	0.34*	–	0.62	0.10	0.57	0.41
(4) FDI	0.48	0.42	0.62	–	0.37	0.40	0.29
(5) Mental capacity	0.32	0.32*	0.10	0.37	–	0.37	0.21
(6) Cognitive level	0.38	0.47	0.57	0.40	0.37	–	0.44
(7) Achievement	0.41	0.40	0.41	0.29	0.21	0.44	-

Dependent variable = achievement:

Independent variable	Multiple r	Multiple r–squared	Beta
General ability	0.410	0.168	0.125
Verbal ability	0.434	0.189	0.149
Spatial ability	0.497	0.247	0.206
FDI	0.498	0.248	–0.062
Mental capacity	0.509	0.259	0.046
Cognitive ability	0.530	0.281	0.213

3.4 The nature of prior knowledge

In contrast to those variables that were the subject of the preceding analysis, another set of interest can be characterized as acquired characteristics. These are most commonly associated with schooling, but it is entirely possible that they might be learned elsewhere.

Interest in such background, or prior knowledge, variables has been generated recently by the research into the development of expertise (Ericsson & Smith, 1991). Of particular relevance to this issue was the contention by Chase and Simon (1973) that the major difference between experts and novices is in their access to relevant domain–specific knowledge.

Relevant prior knowledge is more easily defined in some fields than in others. In the case of chess, used by Chase and Simon, experts were able to recognize on sight approximately 50,000 chess positions. This is similar to the number of different words that a competent reader of the English language might be able to recognize. However, often also included within this group of acquired knowledge bases are information processing, problem solving, or meta–cognitive strategies that are not thought of as psychologically innate (Ericsson & Smith, 1991).

Only recently have science educators begun to include background knowledge as a relevant variable in their studies of achievement. Three categories of measure have emerged during this study. The first is scientific reasoning, the second quantitative reasoning, and the last is content and conceptual knowledge.

The more familiar measures of scientific reasoning skill are process measures such as the Test of Integrated Process Skills (TIPS) or the Process of Biological Investigations Test. However, they are also very similar in many respects to the most commonly used measures of cognitive level, such as the Lawson Test of Formal Operations or the Test of Logical Thinking. Both types of measure have more often been used as dependent than as independent variables in science education research. However, it is at least as reasonable to think of them as measures of generalized background knowledge that would be useful in promoting achievement.

Quantitative reasoning is often represented in research studies as a variable similar to ability or cognitive level, with the implication that it has underlying psychological properties. Indeed, some measures share properties with measures of cognitive level in that they both contain items requiring proportional reasoning. Again, it seems reasonable to think of quantitative reasoning as a type of background knowledge.

Cognitive level, quantitative and scientific reasoning, and scientific achievement are highly inter–correlated. Because of both the psychometric and conceptual similarities between the three variables, a regression was conducted to assess the relative variance shared between them in the prediction equation (Table 6). Achievement was regressed first against cognitive level, and then quantitative reasoning was entered. This step increased the explained variance in achievement from 19 % to 22 %. Scientific reasoning was entered in the last step, increasing the explained variance to 27 %. The Beta for scientific reasoning (.31) is substantially larger than that for cognitive level (.22). The very low Beta for quantitative reasoning (0.09) implies that this variable contributes little to scientific achievement when the variance it shares with the other two variables has been taken into account.

Table 6: Regression of achievement against cognitive level, quantitative reasoning and scientific reasoning

	(1)	(2)	(3)	(4)
(1) Cognitive level	–	0.50	0.57	0.48
(2) Quantitative reasoning	0.50	–	0.49	0.35
(3) Scientific reasoning	0.57	0.49	–	0.48
(4) Achievement	0.44	0.35	0.42	–

Dependent variable = achievement:

Independent Variable	Multiple r	Multiple r–squared	Beta
Cognitive level	0.440	0.194	0.218
Quantitative reasoning	0.465	0.216	0.088
Scientific reasoning	0.521	0.271	0.313

Four types of knowledge measures were identified in this study. The first are standardized assessments of achievement in science, such as the California Achievement Test or College Board examinations. The second are pre–tests, sometimes taken from item banks and often similar or identical to the post–test used in the same study. Third are the number or type of misconceptions held by students. Finally are prior course–taking and achievement, as for example the number of previous science courses and the science grade–point average.

ERIC

There are serious issues regarding the use of background knowledge as a variable in studying achievement. First, these measures tend to share a large amount of variance with the dependent variable, and in some extreme cases are identical. Thus, regressing post–test against pre–test scores enters a kind of circular logic into the research design that is difficult to avoid. The procedure removes from the regression equation a very large amount of the variance in the dependent variable at the first step, and thus reduces the explanatory power of other variables. It also tends to produce ceiling effects for students of high ability, and leads to strange statistical artifacts. An example occurred in a study by Lawson and Worsnop (1992), who regressed gain–scores on pre–test scores and obtained a relatively large negative correlation. This resulted from the fact that students with high background knowledge scores had little room for further growth in achievement, while those that were initially relatively uninformed about the subject showed the anticipated gains. For these reasons, the analysis by means of multivariate statistics of the role of prior knowledge in achievement has been approached with some caution.

The approach taken here, especially in light of evidence from the previous analyses of the independence of scientific reasoning as a factor, is to test the strength of the relationship of scientific reasoning skills to content knowledge and thus to achievement in science. Since no data were available for the relationship between scientific reasoning and content knowledge, the correlation of 0.40 with achievement was substituted. In this analysis (Table 7), achievement was first regressed against scientific reasoning and then content knowledge, which is the anticipated path of this relationship. In this analysis, 16 % of the variance in achievement is explained at the first step, and 22 % at the second. When the order of entry is reversed (not shown), virtually the same result is obtained, with 15 % of the variance in achievement explained at the first step and 22 % at the second. This suggests that the two variables contribute approximately equal and relatively independent variance to this equation. The almost equal value of Beta for the two independent variables supports this conclusion.

Table 7: Regression of achievement against scientific reasoning and prior knowledge (* see text)

	(1)	(2)	(3)
(1) Scientific reasoning	–	0.40	0.40
(2) Prior knowledge	0.40	–	0.39*
(4) Achievement	0.40	0.39*	–

Dependent variable = achievement:

Independent Variable	Multiple r	Multiple r–squared	Beta
Scientific reasoning	0.400	0.160	0.291
Prior knowledge	0.472	0.223	0.274

ERIC

Table 8: Mean correlations between attitude toward science and background variables

Attitude toward science	Mean correlation	Standard deviation	Number of studies
General ability	0.26	0.01	2
Verbal ability	0.10	0.20	2
Cognitive level	0.23	–	1
FDI	0.36	0.10	2
Locus of control	−0.08	0.25	2
Quantitative reasoning	0.21	0.19	3
Scientific reasoning	0.14	0.10	3
SES	0.07	0.02	2
Achievement	0.31	0.14	14

3.5 Attitude

Most authors who have worked with attitude have found its relationship with achievement to be low. Again, the correlations reported by Steinkamp and Maehr (1983) and in this study were similar, and if averaged would yield a value of 0.25.

The relationship between attitude and achievement is a puzzle. It is equally plausible to suggest that attitude causes achievement as that achievement causes attitude. In fact, the position taken here is that there is no necessity of, or evidence for, a causal link between the two.

Taking attitude toward science as a dependent variable, and inquiring separately into its origins leads to a consideration of the nature of the correlations between all measured variables and attitude. A complete matrix of correlations among nine variables has been compiled. Neither the size of this matrix nor available theories about the origins of attitude seem adequate to a complete analyses such has been conducted for achievement. Instead, single regression of attitude against all variables was conducted (Table 9).

Table 9: Regression of attitude toward science against all variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) General ability	–	0.74	0.38	0.48	0.04	0.55	0.54	0.41	0.26
(2) Verbal ability	0.74	–	0.47	0.42	0.19	0.54	0.48	0.40	0.10
(3) Cognitive level	0.38	0.47	–	0.40	0.21	0.50	0.57	0.44	0.23
(4) FDI	0.48	0.42	0.40	–	0.13	0.51	0.47	0.29	0.36
(5) Locus of control	0.04	0.19	0.21	0.13	–	0.06	0.08	0.21	−0.08
(6) Quantitative reasoning	0.55	0.54	0.50	0.51	0.06	–	0.49	0.35	0.21
(7) Scientific reasoning	0.54	0.48	0.57	0.47	0.08	0.49	–	0.40	0.14
(8) Achievement	0.41	0.40	0.44	0.29	0.21	0.35	0.40	–	0.31
(9) Attitude	0.26	0.10	0.23	0.36	−0.08	0.21	0.14	0.31	–

Continuation on the next page

Table 9 (continuation)

Dependent variable = achievement:			
Independent variable	Multiple r	Multiple r-squared	Beta
General ability	0.260	0.068	0.264
Verbal ability	0.294	0.087	−0.283
Cognitive level	0.352	0.124	0.173
FDI	0.428	0.184	0.334
Locus of control	0.441	0.195	−0.154
Quantitative reasoning	0.442	0.195	−0.022
Scientific reasoning	0.460	0.212	−0.205
Achievement	0.511	0.261	0.264

The largest increases in a explained variance in attitude in this solution are associated with the entry of general ability, field–dependence/independence and achievement. The importance of the relationship between these three variables and attitude is also indicated by their relatively high values of Beta.

3.6 Socio–economic status

The correlation between Socio–economic status (SES) and achievement is relatively low, both as reported by Fleming and Malone (1983) and in the reports compiled for this study. An average value for that earlier meta–analysis and this one would be 0.29. Unfortunately, the full matrix of correlation coefficients that would allow tests of this relationship as conducted in previous analyses is not available.

4. Discussion

This was initiated as a feasibility study for the use of meta–analysis in examining the relative impact of a large number of factors on achievement in science. Despite several inadequacies in its current form, it does provide a platform for the discussion of both methodological and substantive issues.

4.1 Methodology

Prior studies (Fleming & Malone, 1983; Steinkamp & Maehr, 1983; Willson, 1983) have tended to focus primarily on the correlations between a variety of variables and attitude or achievement. Where all sets of studies have aggregated a substantial number of primary values for the coefficient of correlation, their results are relatively similar to those presented here. The differences arise mainly in those cases where means are based on small samples. This is very much a problem with some of the analyses reported here. The reader should view the substantive results with caution in those causes where correlation matrices contain values that are based on small samples or are estimated.

The earlier studies have also demonstrated that the magnitude of relationships may vary by age or grade level of the subject and by content area. This has not been possible in the current study. It has also not been possible to meet the full methodological rigor advocated by Hedges (1986). An important objective of any expansion of this study would have to offer more careful and convincing proof that the groupings and categories of measure and study are not biasing the result.

However, in one important respect this study has demonstrated the possibility of meeting one of Hedges' objectives. He and others have criticized meta-analysis for its tendency to over-generalize variable categories, and thus not distinguish among experimental designs, samples or measures that are quite different. This study has demonstrated the feasibility of aggregating sufficient data to actually test hypotheses about specific measures, such as the Raven Progressive Matrices and the Embedded Figures Test.

It has also been demonstrated that it is possible to collect from the literature a sufficient number of correlation pairs to engage in a secondary analysis using multivariate statistical procedures. Of course, all appropriate caution should also be observed with regard to the interpretation of these analyses. It is certainly not true that an observed relationship between two variables proves a causal link. The use, wherever possible, of the hypothesis testing procedures advocated by Kerlinger (1973, pg. 628) seems to meet the normal objections to some degree. However, the wise reader will remember that the results of a multivariate procedure are biased by assumptions about the relationships between variables, and are no better than the theory that underlies them.

4.2 Results

The data collected for this study proved more useful in the study of the relationship between psychological factors, background knowledge and achievement than it did for any careful analysis of the role of socio-economic status or of attitude toward science. At best, it can be concluded at this time that neither SES nor attitude are strongly related to achievement in science. This is fully consistent with the results of prior research.

Neither the data collected for this study nor current theory were adequate to the task of a detailed analysis of the antecedents of attitude toward science. A routine but relatively uninformative regression of attitude against eight variables indicated that the strongest association was with field-dependence/independence. Achievement and general ability had slightly lower, and equal, values for Beta. No other variable accounted to any large extent for variance in attitude.

This study has re-emphasized the importance of general ability in scientific achievement. Much more significant, though, is the refutation of earlier hypotheses that spatial ability contributes little to scientific achievement once the contribution of general ability has been controlled for.

The importance of spatial ability was again made clear in the analysis of neo-Piagetian factors. The well known relationship of field-dependence/independence to achievement appears to depend almost entirely upon its spatial component. Although not tested in this study, there is reason to think that the Raven Progressive Matrices Test also has a strong spatial aspect.

While this analysis does not prove causal links, there is corroborating research that is very suggestive. Although spatial skills are rarely part of the curriculum, they can be quite successfully taught (Lord, 1985, 1987), and such instruction has been shown to improve conservation task performance for young children (Dolecki, 1981) and physics achievement for college students (Pallrand & Seeber, 1984). There is very definite reason to believe that better spatial skills result in improved performance in science, and that the results can be educationally meaningful.

This study has shown that the neo-Piagetian variable of memory capacity is also contained within the Raven Progressive Matrices. Studies of expertise have demonstrated the importance of this ability in performance as diverse as that of bartenders and chess masters. At first, these experts seemed to violate the rule (Miller, 1956) that strings of information more than nine units long are almost impossible to remember. However, subsequent studies revealed that experts were using chunking strategies that were unavailable to novices and that allowed them to remember substantially more.

As was the case for spatial ability, there is ample research suggesting that mnemonic skills can be taught (Pressley, Levin & Delaney), and that this leads to important increases in achievement in science (Banks & Piburn, 1986). And as was true in the earlier instance, this seems to provide evidence for a causal link from memory capacity to scientific achievement.

The question of whether cognitive level is an independent factor in scientific achievement has not been fully resolved. Its Beta in a regression of achievement against all psychological variables is the largest of any, but the increase in shared variance with achievement at its entry into the equation is only 2 %.

While quantitative ability does not appear to be a major factor in scientific achievement, both scientific reasoning and background knowledge are. Furthermore, these latter two variables are relatively independent of one another, and share approximately equal amounts of variance with scientific achievement after the effects of psychological variables have been controlled.

These results allow a general model for the development of scientific achievement that is not unlike the contemporary view of expertise. General intelligence, although a factor, is surprisingly unimpressive as a predictor of scientific achievement, as it is of other forms of outstanding performance. In fact, Ericsson and Smith state that IQ tests "have been remarkably unsuccessful in accounting for individual differences in levels of performance in the arts and sciences and advanced professions" (1991, pg. 5). On the other hand, both memory capacity and background knowledge are major components of almost all views of expert performance. The current controversy, well worth exploring in the case of achievement in science, is whether memory strategies and ability are general or context-specific (Peverly, 1991).

There is less discussion in the literature on expertise about the role of spatial ability. The exception seems to be chess, where "superior spatial ability often is assumed to be essential" (Ericsson & Smith, pg. 6). One set of results linking chess masters to superior ability in memory tests involving the position of chess pieces indicated that a factor in their performance might be superior visual memory. Although the relationship of spatial ability to achievement in science seems clear in this study, the mechanisms by which this ability is utilized by superior achievers are not clear at all, and should be a fruitful subject of further study.

ERIC

While the literature on expert–novice performance contains most of the factors that have been identified in this study as important to scientific achievement, it is by no means limited to those. While this is not the place for a further discussion of this literature, it has the potential to be a rich mine for research in science education.

Finally, the utility of meta-analytic techniques for the aggregation of data from previous studies has been demonstrated. It would have been quite difficult to have conducted a single study that would have combined the power and number of variables that are represented here. Correlation coefficients are routinely reported in studies of all types, and can be used by means of this technique for the testing of theoretical questions.

References

Banks, G. & Piburn, M. D. (1986). *The use of mnemonics in teaching middle school science.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Francisco, CA.

Beilin, H. (1987). *Current trends in cognitive development research: Towards a new synthesis.* In: B. Inhelder, D. deCaprona & A. Cornu–Wells (Eds.), Piaget today. Hillsdale, NJ: Lawrence Erlbaum.

Bolding, (1985). *Statistics with finesse.* Fayatteville: AR.

Bodner, G. M. & McMillen, T. L. B. (1986). *Cognitive restructuring as an early stage in problem solving.* Journal of Research in Science Teaching, vol. 23 (8), 727 – 738.

Chase, W. G. & Simon, H. A. (1973). *The mind's eye in chess.* Ch. 1 in W.G. Chase (Ed.), Visual information processing. New York: Academic Press.

Dolecki, P. (1981). *The effects of a perceptual training program on conservation task performance and field independence in first grade children.* Unpublished doctoral dissertation, Rutgers University, NJ.

Ekstrom, R. B., French, J. W., Harman, H. H. & Dermen, D. (1976). *Manual for kit of factor–referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Ericsson, K. A. & Smith, J. (1991). *Prospects and limits of the empirical study of expertise: an introduction.* In: Ericsson & Smith (Eds.), Toward a general theory of expertise: Prospects and limits. Cambridge: Cambridge University Press.

Fleming, M. L. & Malone, M. R. (1983). *The relationship of student characteristics and student performance in science as viewed by meta–analysis research.* Journal of Research in Science Teaching, vol. 20 (5), 481 – 495.

Glass, G. V. (1976). *Primary, secondary, and meta–analysis of research.* Educational Researcher, Nov. 3 – 8.

Glass, G. V., McGaw, B. & Smith, M. L. (1981). *Meta–analysis in social research,* Calif: Sage Publications.

Hedges, L. V. (1986). *Issues in meta–analysis.* In: E. Z. Rothkopf (Ed.), Review of Research in Education, v. 13. Washington, DC: AERA.

Hedges, L. & Olkin, I. (1985). *Statistical methods for meta–analysis.* New York: Academic Press.

Hedges, L. V., Shymansky, J. A. & Woodworth, G. (1989). *Modern methods of meta–analysis.* Washington, DC: National Science Teachers Association.

Kerlinger, F. N. (1973). *Foundations of behavioral research.* New York: Holt, Rinehart & Winston.

ERIC

Lawson, A. (1985). *A review of research on formal reasoning and science teaching*. Journal of Research in Science Teaching, vol. 22 (7), 569 – 618.

Lohman, D. F., (1988) *Spatial abilities as traits, processes and knowledge*. Ch 6 in R. J. Sternberg (Ed.), Advances in the psychology of human intelligence, v. 4, Hillsdale, NJ: Lawrence Erlbaum Associates.

Lord, T. R. (1985). *Enhancing the visuo-spatial aptitude of students*. Journal of Research in Science Teaching, vol. 22 (5), 395 – 405.

Lord, T. R. (1987). *Spatial teaching*. The Science Teacher, February, 32 – 34.

Miller, G. A. (1956). *The magical number seven, plus or minus two*. Psychological Review, vol. 63, 81 – 97.

Nie, N. H., Hull, C. H., Jenkins J. G., Steinbrenner, K. & Bent, D. H. (1975). *SPSS: Statistical package for the social sciences*, second edition. New York: McGraw-Hill.

Pallrand, G. & Seeber, F. (1984). *Spatial ability and achievement in introductory physics*. Journal of Research in Science Teaching, vol. 21 (5), 507 – 516.

Pascual-Leone, J. (1969). *Cognitive development and cognitive style: A general psychological integration*. Unpublished doctoral dissertation, University of Geneva.

Peverly, S. T. (1991). *Problems with the knowledge-based explanation of memory and development*. Review of Educational Research, vol. 61 (1), 71 – 93.

Pressley, M., Levin, J. R. & Delaney, H. D. (1982). *The mnemonic keyword method*. Review of Educational Research. vol. 52 (1), 61 – 91.

Schmid, J. E., Koch, G. G. & LaVange, L. M. (1991). *An overview of statistical issues and methods of meta-analysis*. Journal of Biopharmaceutical Statistics, vol. 1 (1), 103 – 120.

Sperry, R. W. (1961). *Cerebral organization and behavior*. Science, 133, 1749 – 57.

Steinkamp, M. W. & Maehr, M. L. (1983). *Affect, ability, and science achievement: A quantitative synthesis of correlational research*. Review of Educational Research, 53 (3), 369 – 396.

Sternberg, R. J. (1980). *Representation and process in linear syllogistic reasoning*. Journal of Experimental Psychology: General, 109, 119 – 59.

Sternberg, R. J. (1990). *Metaphors of mind: C[illegible]ptions of the nature of intelligence*. Cambridge: Cambridge University Press.

Willson, V. L. (1983). *A meta-analysis of the relationship between science achievement and science attitude: kindergarten through college*. Journal of Research in Science Teaching, vol. 20 (9), 839 – 850.

Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. R. & Karp, S. A. (1962). *Psychological differentiation: studies of development*. New York: Wiley.

COMMENTARY OF THE DISCUSSANT

Elisabeth Schach, Universität Dortmund, Germany

The following comments are based on a review of the statistical aspects of meta-analysis as discussed by Schmid, Koch, and LaVange (1991). The authors present advantages, critical aspects and remedies of how to handle statistical measures in meta-analytical procedures. My comments consist of presenting selected methodological aspects of such meta–analyses and of applying them to Dr. Piburn's presentation.

Table 1 shows aspects of meta–analysis as given by Schmid, Koch, and LaVange in their 1991 review paper. The obvious advantages of these analyses consist of the reliability of estimates and the credibility of empirical results that are derived from large (sample size) studies. The associated disadvantage frequently is that researchers performing meta–analyses are unable to critically review the methodology of all relevant studies and to consciously select the appropriate ones on the basis of this review's results. This is due to the fact that studies that are included in meta–analytical summaries are usually extracted from scientific journals, where only limited space is available for the documentation of study methods. Schmid, Koch, and LaVange suggest remedies of how to improve the results of meta–analyses. However, some of these improvements are only feasible if the documentation of methods of the respective studies is reasonably uniform and complete. Thus, in order to improve the applicability of meta–analytical procedures, journal editors should be encouraged to grant paper authors sufficient space for the description of study methods. Due to this flaw, the potential of meta–analysis is far from having been exhausted.

Table 1 shows selected aspects of meta–analyses (advantages, problems, and remedies). As Pigeot discusses these in this volume (1992) with respect to aggregating correlation coefficients, no further comments are presented here.

Table 1: Methodological aspects of meta–analyses (according to Schmid, Koch, and LaVange 1991)

Advantages	Large sample sizes Increase in quality of information
Problems	Investigators' lack of control of data Publication bias Unobservable study aberrations Incompletely reported data Nonindependence of studies Nonindependence of subjects Heterogeneity of studies
Remedies	Thorough search for relevant studies Well–defined criteria for study inclusion Weighting of studies Analysis update, as more studies become available Stratification of studies

Table 2 shows the results of the application of Schmid, Koch, and LaVange's methodological criteria to Dr. Piburn's paper. The application of any of their criteria is possible due to Dr. Piburn's excellent documentation of study characteristics as far as they were available to him. Piburn's contribution demonstrates in part how one could proceed to collect the necessary data. However, I believe, that a more detailed description of study methods (of the original studies) would be required in order to be able to apply Schmid et al's methodological criteria for study selection and handling in a specific meta–analysis. The check list provided by them may serve the purpose to accomplish this eventually.

Table 2: Application of Table 1 criteria to meta-analysis by Dr. Piburn

Advantages	Sample sizes of studies not given (small?) Increase in quality of information (some of it seems to be conflicting)
Problems	Investigators' lack of control of data (not discussed in detail) Publication bias (not discussed) Unobservable study aberrations (not discussed) Incompletely reported data (not discussed) Nonindependence of studies (not discussed) Nonindependence of subjects (not discussed) Heterogeneity of studies (discussed)
Remedies	Thorough search for relevant studies Well-defined criteria for study inclusion (might be more specifically stated) Weighting of studies (not done) Analysis update as more studies become available (reported analysis was retrospective) Stratification of studies (not done)

A few more specific comments are in order. Sample sizes of the included studies are not presented. This should be done in a future report. Comments with respect to conflicting results of reported studies (as given) are important, as they show that the aggregation of evidence may support several hypotheses instead of just one. It is suggested that the problem areas of meta-analyses of the specific studies be discussed and then evaluated with respect to their impact on meta-analytical results.

Some of the remedies suggested by Schmid et al may be applicable in this meta-analytical study as well. They are: search for relevant studies in more than one journal, precise statement of criteria for study inclusion, weighting of study correlation coefficients on the basis of sample sizes of original studies, and stratification of studies on the basis of common variables. Even though all of these suggestions involve an even more complicated process of study documentation and selection, there may be rewards for doing this. The result will be a smaller set of more homogeneous investigations, whose aggregated results provide what we expect of meta-analyses — a more reliable statement about postulated relationships about attitude and achievement in science.

A few more comments relating to study methods may be added. All of them relate to specific methodological aspects of the selected studies. When means of correlation coefficients are computed, it might be reasonable to ascertain the appropriateness of these in original studies. It would also be necessary to examine which type of variable they were applied to (ordinal, interval). Pearson product-moment correlation coefficients are appropriate for linear relationships, even though it might be very difficult to examine that precondition of the measure. This point holds true for multiple R^2 of the original studies as well. In order to be able to use variables in aggregate analyses, it would be important to understand which criteria were applied for the inclusion of variables in reported linear regression analyses. Were these inclusions based on the reported hypotheses or were items selected on the basis of prior analyses aimed at examining their impact on dependent variables. A further question relates to whether correlation coefficients might have been biased to start with (due to the fact that they were calculated on the basis of aggregate data).

Furthermore, the question which outcome variables tap which dimensions of science achievement needs to be answered at some point. Results of this examination might provide guidance for the selection of outcome variables in future meta-analyses.

A final question is whether meta-analysis results would have been different had study weighting by sample size or stratification of homogeneous studies been applied.

Despite of these critical remarks, it was a pleasure to review the paper by Dr. Piburn, because this review provided the commentator the opportunity to examine selected methodological aspects of meta-analyses and to review their impact on empirical results from such studies. Meta-analyses seems to be an important tool for empirical research. As in all applied research, the relevance of its results depend largely on available data. Thus, the meta-analysis researcher needs to critically review studies that are to be considered for such summaries. Readers of meta-analysis results need to be convinced of the credibility of results derived from such study aggregations.

References

Piburn, M. D. (1992). *Meta-analysis and multivariate procedures for the study of attitude and achievement in science.* (see above)

Pigeot, I. (1992). *Correlation and causality.* (in this volume).

Schmid, J. E., Koch, G. G. & LaVange, L. M. (1991). *An overview of statistical issues and methods of meta-analysis.* Journal of Biopharmaceutical Statistics vol. 1 (1), 103 - 120.

SUMMARY OF THE PLENARY DISCUSSION

Jutta Theißen, Universität Dortmund, Germany

Meta-analysis seems to be a reasonable technique, however, further discussions are needed to see whether more aspects should be considered in a meta-analysis. This technique is a good idea, but there is not enough information to judge how good exactly it is. It was pointed out that, especially for a meta-analysis, good co-operation between educationalists and statisticians was very important.

The validity of the independent variables was questioned. For example, you do not exactly know what a researcher meant when he/she used the independent variable "memory". You do not know the memory of what he/she meant, and whether the investigator talked about long term or short term memory. However, Michael D. Piburn considered this problem. He categorised the variables like the investigators did; and to get a result of high validity he only took studies with variables that are very common, which seems to be a general tendency, you just see the same data in all studies.

Another item of discussion was the term "achievement". You get information about achievement with the help of certain achievement tests. As most of these tests are multiple choice tests it was doubted whether they were appropriate instruments to measure achievement. Furthermore, there are so many different ideas of the term "achievement" that it is dangerous to put them together. Most measures are too narrow, only one small aspect of achievement is covered.

ERIC

The problem of meta–analyses lies within the studies that are published. Though some tests are very well known, so that you exactly know what the variables were, generally, the tests and variables that are used in the studies are not well–described. Thus the question is how the outcome of a meta–analysis can be interpreted. The quality of a meta–analysis stands and falls with the quality of the studies it refers to, and the meta–analyser cannot do anything about the studies that are published.

ERIC

CORRELATION AND CAUSALITY

Iris Pigeot, Universität Dortmund, Germany

Abstract

In educational research, empirical studies are often conducted to find out the influential variables for the achievement of pupils. Besides the method of education, which is an important variable for this achievement, other possibly influential variables such as sex, attitude toward their teacher or their school, and variables related to their social environment are recorded. In addition, the achievement of the pupils itself has to be measured.

To describe and to measure the degree of association between two of the above variables of interest, correlation coefficients are often calculated. But it has to be taken into account, that a correlation coefficient does not necessarily give an indication for a cause–effect–relationship between two variables. Here the so called path analysis is helpful when being confronted with the problem of giving causal interpretations of observed correlations.

In this paper, an introduction to path analysis is given where technical details are omitted as far as possible. Furthermore, a concept is mentioned which cannot be used to confirm a causal relationship between two variables but to increase the evidence in it by aggregating correlation coefficients from different studies dealing with the same research problem. Such a combination of statistical results across studies is usually called meta–analysis. It is a reanalysis of data which have been collected by other investigators.

1. Introduction

Most empirical studies are conducted to find explanations of e. g. certain events or of a special behaviour. In educational research, for instance, it is often of interest to investigate the influential variables for the achievement of pupils. Usually, the investigator has a certain causal scheme in mind when planning an empirical study. To reject or to confirm this scientific hypothesis concerning the causal structure of interest, he or she tries then to measure at least those variables which are supposed to be involved in the cause–effect–relationship and to collect data accordingly. Being interested in the influential variables for pupils' achievement, for instance, the method of education as a considerably influential factor and other possibly influential variables such as sex, attitude toward teacher or toward school, profession and social status of parents or other variables concerning the pupils' social environment are typically recorded. In addition, the achievement of the pupils itself has to be measured. To evaluate the strength of the assumed association between, say, the pupils' achievement and their attitude toward school, appropriate correlation coefficients are often calculated. But here, a conflict occurs related to the fact that a high observed correlation is not always an indicator of a causal relationship. That means, a high correlation does not necessarily imply that a change in one of the two considered variables induces a change in the second one. If, for instance, a high correlation is observed between e. g. science achievement and parents' profession, it cannot directly be concluded that a change in the parents' profession leads to a change in the achievement of the child.

In this paper, two different concepts are presented which are both helpful when dealing with the problem of causal interpretations of observed correlations. For this purpose, introductions to these concepts are given where technical details are omitted as far as possible. The advantages and limitations of these methods are discussed in the following sections. In addition, their application in the context of interest is demonstrated and illustrated using practical examples.

The first concept to be presented in Section 2 is the so-called path analysis. It deals with finding a causal structure compatible with the collected data. Although especially developed for research in genetics by the population geneticist Sewall Wright in the 1920s, it is nowadays also of practical importance in fields of research such as educational or social sciences.

Section 3 consists of a discussion of the second approach called meta-analysis which is also increasing in importance in educational and social research. This method is a reanalysis of data which have been collected by other investigators. That means, it combines the statistical results derived in former empirical studies dealing with a similar research problem. It has to be pointed out that it cannot confirm a causal relationship between two variables but it can increase the evidence in it by e. g. aggregating correlation coefficients across studies. Thus, a meta-analysis can support a supposed causal relationship if this is observed with a similar intensity in several independent studies.

2. Introduction to path analysis

In this chapter, the basic concept of path analysis is presented. This method dates back to Wright (1921, 1934) and it can be regarded "as a flexible means of relating the correlation coefficients between variables in a multiple system to the functional relations among them" (Wright, 1934). The purpose of this chapter is not to give a detailed description of all methodological, statistical, or technical aspects related to a path analysis, but to give an idea how this method works. Therefore, the application of this method is illustrated using a simple model. For further details on this method see for instance Land (1969), Heise (1969), Blalock (1971, especially Part II), and Duncan (1975). An elementary introduction to path analysis can be found in Li (1977). The exposition of Kang and Seneta (1980) gives a representation of this technique where a description of the underlying statistical structure is emphasized.

2.1 The basic concept of path analysis

In empirical studies in educational or social sciences, there is usually a large number of variables which could influence the outcome variable of interest. This is also the case, for

ERIC

instance, in physical sciences, but investigators in educational or social research are especially confronted with the problem that typically they cannot carry out experiments where step by step variables are controlled or eliminated as possibly influential variables. In such fields of research, all variables have to be considered simultaneously without the possibility of changing only one variable while keeping the remaining variables unchanged. Here, the so-called path analysis offers a possibility to verify or to reject a postulated scheme of causal relationships between pairs of variables considered in an empirical study. Conducting a path analysis can roughly be divided in three steps.

In a first step, the investigator has to fix a model which contains all variables of which it is assumed that they are responsible for the outcome variable, e. g. students' achievement.

In a second step, these variables have to be ordered with regard to a certain causal scheme which the investigator has in mind and which results e. g. from theoretical considerations, from former studies with a comparable research question, especially from observed (partial) correlation coefficients, or from a given chronological order. In this step, a pictorial representation called path diagram is a useful instrument to illustrate correlations or hypothetically causal relationships between each pair of variables by different types of lines. This diagram starts with drawing arrows from each direct 'cause' to each 'effect'.

In the last step, this hypothetical model has to be checked using statistical methods. Among others the so-called path coefficients are usually calculated. These measures are essentially based on correlation coefficients and indicate the strength of a postulated causal relationship. At this point it is not unusual that the statistical analysis calls for a revision of the hypothetical model. Consequently, a path analysis has to be carried out again now based on the modified model.

The last problem results from the fact that there are quite a number of possible arrangements of several variables to be included in a path analysis. Thus, the achieved results concerning the relationships among these variables cannot be regarded in any absolute sense. Each arrangement is connected with a special causal scheme which the investigator has in mind at the beginning of the analysis. Therefore, the gained results describe the relationships just from the particular investigator's point of view.

Obviously, it seems unavoidable that several causal structures have to be checked until the results are indeed plausible. This also implies, as pointed out by Li (1977, pp. 165 – 166), that "... it is the responsibility of the investigator to choose one or more [path diagrams] (preferably more) to serve as starting points for analysis and interpretation. The investigator should always be ready for new diagrams before he obtains a consistent and reasonable interpretation of the causal system within the limit of his data."

Summarizing the basic idea of this method, "it appears to be, in essence, a technique for measuring the strength of postulated causal relationships, and substantiating (or rejecting) the internal 'causal' consistency of a network of such relationships" (Kang, Seneta, 1980, p. 218). The path diagram and the path coefficients can be considered as the essential components of a path analysis.

ERIC

2.2 The technique

A path analysis is closely related to a linear regression analysis as well as to a correlation analysis. It also analyzes the linear relationship of so-called 'endogenous' (dependent) and 'exogenous' (independent) variables, where exogenous variables are considered as having no direct causes. All variables have to be standardized to mean zero and variance unity before conducting a path analysis. Especially, when we have only one variable Y dependent on, say, $X_1,\dots,X_k$, the path analysis and the linear regression lead to the same results. However, a path analysis is more than a simple regression, it allows the treatment of a complete network of variables involving more than one equation. Thus, a path analytic model can be regarded as a set of structural equations which represents the postulated causal connections among the variables involved in the investigation. There are two types of path analytic models to be distinguished. In so-called recursive models, no two variables can be simultaneously cause and effect of each other and no variable can be an indirect cause of itself, otherwise it is said to be nonrecursive. Here, a variable X is said to be an indirect cause of another variable, say, Z, if X affects Z through a chain of other variables without any path directly connecting X and Z.

As already mentioned above, it is customary to represent the postulated causal scheme in a path diagram, where initially arrows are drawn from each direct cause to each endogenous variable. Then, for each dependent variable a residual variable is added standing for the aggregation of all outside disturbances influencing the dependent variable. The inclusion of these variables is necessary, since not all of the variance of each dependent variable can be explained by the identified affecting variables. Thus, the residual variables represent all other possible sources of variation in the endogenous variables. The residual variable is, therefore, itself regarded as an independent variable and as an direct cause. At last, exogenous variables which are known to be correlated, but not causal related are connected by a curved double-headed arrow. In such a path diagram, each single-headed arrow is finally marked with the corresponding path coefficient and each curved line is analogously marked with the corresponding correlation coefficient. A more detailed description of the rules for constructing a path diagram can be found for instance in Li (1977, pp. 106 – 121, pp. 161 – 165) and in Kang, Seneta (1980, pp. 222 – 229).

In the remaining part of this paragraph, the technique of a path analysis will be introduced using a model with only two variables. A more general approach will be added.

At first, the assumptions of a path analysis are summarized following the representation in Heise (1969). Here, the path analytic model is assumed to be recursive.

1. The path analytic model consists of structural equations which are all linear, i.e. all dependent variables are treated as linear functions of the independent and the residual variables. If this assumption is not fulfilled, the interpretation of estimated path coefficients as well as of observed correlations can be misleading. Thus, a small value (close to zero) does not necessarily mean, that the considered pair of variables is indeed uncorrelated. There can exist a highly non-linear functional relation between those variables.
2. It is necessary that all variables affecting the outcome variable are specified and included in the model. For this purpose, the investigator has to clearly distinguish between the independent variables (input) and the dependent variables (output).

Furthermore, a causal structure has to be postulated. This implies that the dependent variables must be ordered in terms of their causal priorities.

3. Moreover, it is often assumed that the residual variables are uncorrelated with each other or with the independent variables. This assumption is closely related to the so-called identification problem which can occur in the estimation of path coefficients (c. f. Heise, 1969, pp. 52 – 57).
4. There are no measurement errors, i.e. the used instruments to measure the variables of interest must have high reliability and validity.
5. Since the estimation of path coefficients is based on least squares or other regression techniques, the typical assumptions of a multiple regression have to be met, too. These assumptions will not be listed here (for a brief summary see e. g. Heise, 1969, p. 57) except the problem of multicollinearity. Although it is in general allowed that the exogenous variables are correlated, these correlations should not be very large, because then the effects of the exogenous variables can hardly be separated from each other.

Under the first assumption, the dependence of a single endogenous variable X_1 on k−1 exogenous variables $X_2, \ldots, X_k$ and on a residual variable $\tilde{U}$ can be described by the following linear equation:

$$X_1 = b_{12}X_2 + \ldots + b_{1k}X_k + b_{1\tilde{U}}\tilde{U}$$

b_{1j}, $j = 2,\ldots,k$, $b_{1\tilde{U}} \in \mathbb{R}$. Let us assume (without loss of generality) that $E(X_j) = 0$, $j = 1, \ldots, k$, and $E(\tilde{U}) = 0$, where E(X) denotes the mean of a random variable (r.v.) X.

Thus, it only remains to standardize the variables to variance unity. Dividing these terms by their standard deviations

$$\sigma_j = \sqrt{\sigma_j^2} = \sqrt{Var(X_j)},$$

where Var(X) denotes the variance of a r.v. X, yields:

$$Z_1 = b_{12}\frac{\sigma_2}{\sigma_1}Z_2 + \ldots + b_{1k}\frac{\sigma_k}{\sigma_1}Z_k + b_{1\tilde{U}}\frac{\sigma_{1\tilde{U}}}{\sigma_1}U,$$

where

$$Z_j = \frac{X_j}{\sigma_j} \quad \text{and} \quad U = \frac{\tilde{U}}{\sigma_{\tilde{U}}} \quad \text{with} \quad \sigma_{\tilde{U}} = \sqrt{Var(\tilde{U})}.$$

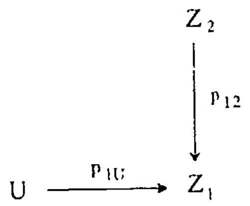
The path coefficients p_{1j}, $j = 2, \ldots, k$, and p_{1U} are now defined as $p_{1j} = b_{1j}(\sigma_j/\sigma_1)$ and $p_{1U} = b_{1\tilde{U}}(\sigma_{\tilde{U}}/\sigma_1)$, where the first subscript denotes the dependent and the second subscript the independent variable. It can be seen from this representation that path coefficients are of the same type as standardized regression coefficients. It should be noticed that path coefficients can take values greater than unity and less than −1, which is in contrast to correlation coefficients.

For illustrative purposes, let us first restrict to the case where we have only one endogenous and one exogenous variable. Then, we obtain the following path analytic model:

(1) $Z_1 = p_{12}Z_2 + p_{1U}U$

with the corresponding path diagram

ERIC



The correlation coefficient ϱ_{12} of two r.v.s Z_1 and Z_2 is defined as

$$\varrho_{12} = \frac{Cov(Z_1,Z_2)}{\sqrt{Var(Z_1)}\sqrt{Var(Z_2)}},$$

where

$$Cov(Z_1, Z_2) = E(Z_1 Z_2) - E(Z_1)E(Z_2)$$

denotes the covariance of Z_1 and Z_2 (see also Chapter 3). In contrast to a path coefficient, the subscripts of a correlation coefficient are not ordered, i. e. $\varrho_{12} = \varrho_{21}$.

Since Z_1, Z_2 are standardized variables with

$$E(Z_i) = 0 \quad \text{and} \quad Var(Z_i) = 1, \quad i = 1,2$$

it follows:

(2) $$\varrho_{12} = E(Z_1 Z_2).$$

Now let us multiply Equation (1) with Z_2, which results in

(3) $$Z_2 Z_1 = p_{12} Z_2^2 + p_{1U} Z_2 U.$$

Taking the expectation yields:

(4) $$E(Z_2 Z_1) = p_{12} E(Z_2^2) + p_{1U} E(Z_2 U).$$

Because of the third assumption and Equation (2), Equation (4) is equivalent to

$$\varrho_{12} = p_{12} 1 + p_{1U} 0.$$

That means, in the above Model (1), the path coefficient is just the correlation coefficient and can therefore directly be estimated by the empirical correlation coefficient r_{12} (see also Chapter 3). Analogously, it can be shown that $\varrho_{1U} = p_{1U}$. Moreover, it can be seen from Equation (1) by multiplying with Z_1 and taking the expectation, that

$$E(Z_1^2) = p_{12} E(Z_2 Z_1) + p_{1U} E(Z_1 U).$$

This implies

$$\varrho_{11} = 1 = p_{12}^2 + p_{1U}^2$$

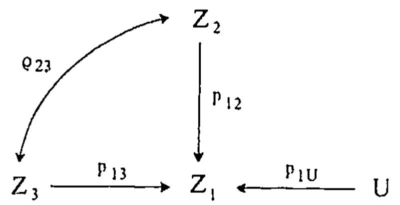
or equivalently

(5) $$p_{1U} = \sqrt{1 - p_{12}^2}.$$

Expanding the above procedure to a model with two exogenous variables, i.e.

(6) $$Z_1 = p_{12} Z_2 + p_{13} Z_3 + p_{1U} U$$

with the path diagram



yields the following equations for the path coefficients

$$\varrho_{12} = p_{12} + p_{13}\varrho_{23}$$

$$\varrho_{13} = p_{12}\varrho_{23} + p_{13} \qquad (7)$$

$$\varrho_{11} = 1 = p_{12}\varrho_{12} + p_{13}\varrho_{13} + p_{1U}^2$$

$$p_{1U} = \sqrt{1 - p_{12}\varrho_{12} - p_{13}\varrho_{13}}\ .$$

Solving the first two equations of (7) with respect to p_{12} and p_{13}, we get

$$(8) \qquad p_{12} = \frac{\varrho_{12} - \varrho_{13}\varrho_{23}}{1 - \varrho_{23}^2} \quad \text{and} \quad p_{13} = \frac{\varrho_{13} - \varrho_{12}\varrho_{23}}{1 - \varrho_{23}^2}\ ,$$

which can be estimated by

$$(9) \qquad \hat{p}_{12} = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} \quad \text{and} \quad \hat{p}_{13} = \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2}\ .$$

It can be seen from (9), that $\hat{p}_{12}$ and $\hat{p}_{13}$ are identical with the corresponding least square estimators of the standardized partial regression coefficients.

Coming back to Equation (7) or explicitly to

$$\varrho_{12} = p_{12} + p_{13}\varrho_{23}$$

$$\varrho_{13} = p_{12}\varrho_{23} + p_{13},$$

it should be mentioned, that the correlation of a dependent and an independent variable can obviously be decomposed in a so-called direct effect represented by the path coefficient and in a so-called indirect effect via the product of the correlation coefficient between the two independent variables and the path coefficient of the second independent variable. Thus, the indirect effect of e. g. Z_2 on Z_1 is given as $\varrho_{12} - p_{12}$ and can also be estimated using the observed correlation coefficients.

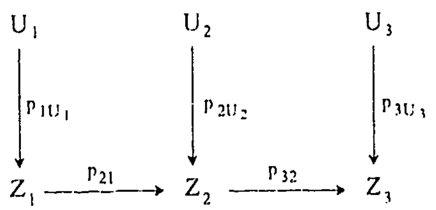
Model equation (6) and the resulting equations for the path coefficients can directly be expanded for more than two exogenous variables. The corresponding formulae can be found for instance in Land (1969, p. 20). The case of more than one endogenous variable will be discussed only for a model with three dependent variables (c. f. Land, 1969, pp. 29 – 32). Especially, the following model and the corresponding path diagram are postulated:

$$Z_1 = p_{1U_1}U_1$$

$$Z_2 = p_{21}Z_1 + p_{2U_2}U_2$$

$$Z_3 = p_{32}Z_2 + p_{3U_3}U_3$$

with



As it can be seen from the model equations as well as from the diagram, the residuals are assumed to be uncorrelated. If this assumption is not fulfilled, their correlations have to be taken into account when estimating the path coefficients (see also Land, 1969, p. 30). Based on the above model, we get the following equations for the path coefficients:

$$p_{21} = \varrho_{21}$$
$$p_{32} = \varrho_{32}$$
$$p_{21}\,p_{32} = \varrho_{13}$$
$$\varrho_{22} = 1 = p_{21}^2 + p_{2U_2}^2$$
$$\varrho_{33} = 1 = p_{32}^2 + p_{3U_3}^2 .$$

A discussion of more complicated models can be neglected in this paper, because no further insight can be gained from such models with respect to the principles of a path analysis.

At the end of this paragraph, only one further aspect should be mentioned. If some of the estimated path coefficients are close to zero, these paths can possibly be deleted from the path diagram. Having deleted apparently nonexisting paths from the model, the remaining path coefficients have to be estimated again based on the new, so-called trimmed model. This procedure is discussed in Heise (1969, pp. 59 – 61).

2.3 An example

The following study is reported here for illustrative purposes. It deals with a path analysis which was conducted to find an explanation of chemistry students' achievement on volumetric analysis problems (Anamuah–Mensah, Erickson, Gaskell, 1987). The authors investigated possibly causal relationships among direct proportional reasoning, inverse proportional reasoning, prerequisite concepts or concepts subsumed by volumetric analysis calculations and performance on volumetric analysis calculations.

For this purpose, the authors selected 402 grade twelve chemistry students from 17 classes in ten schools in British Columbia, Canada. Out of this group, only 265 students from 14 intact classes in eight schools participated fully by writing all the tests shortly described below:

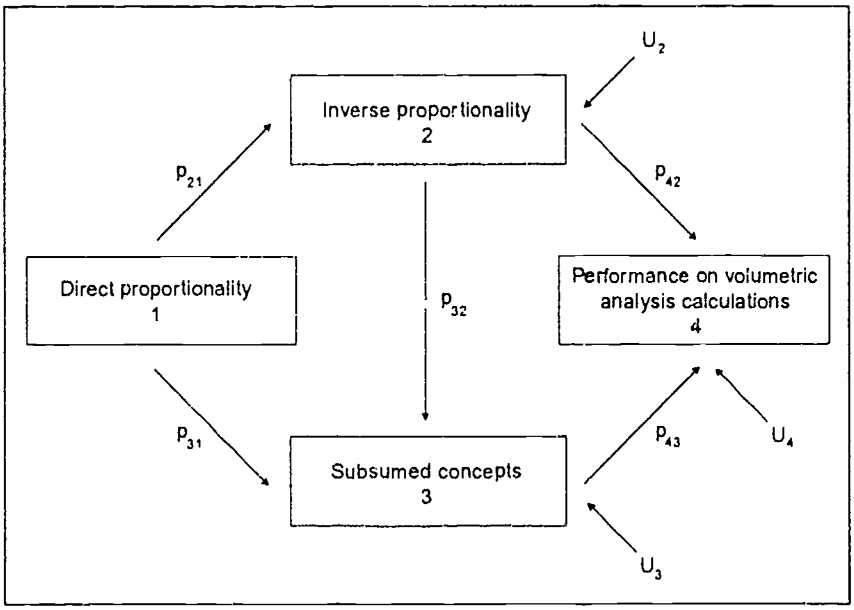
The authors used a 14–items group administered proportionality test to assess proportional reasoning which was composed of a direct proportionality and an inverse proportionality subtest to measure the two different parts of proportional reasoning. Furthermore, knowledge of the prerequisite concepts subsumed by volumetric calculations was measured by a 28–items multiple choice subconcepts test. At last, a 15–items volumetric analysis test was used to measure performance on volumetric analysis problems.

The total study took place in winter, 1980.

Because of former studies the authors distinguished between two groups of students depending on whether the students use algorithms with or without understanding the relations involved. Here, the results are primarily reported for those chemistry students who use algorithms without understanding the underlying relations.

To carry out a path analysis, the investigators first had to propose a causal system based on theoretical considerations and on knowledge resulting from a large number of reported studies dealing primarily with students' understanding of chemical concepts. These studies made use of the proportional reasoning schema of Piaget and the cumulative learning theory of Gagne. Taking all aspects of interest into account, the authors developed the hypothetical model presented in the following path diagram for both groups of students (for a more detailed discussion see Anamuah-Mensah, Erickson, Gaskell, 1987, pp. 726 – 728). In this path diagram, only the connecting paths from each cause to each effect are represented by single-headed arrows. Obviously, it represents a recursive model, because no variable is an indirect cause of itself. As it can be seen from this diagram, direct proportionality (DP) is considered as exogenous variable, while inverse proportionality (IP), subsumed concepts (SC), and performance on volumetric analysis calculations (PVAC) are endogenous variables. The residual variables U_2, U_3, U_4 are connected with the corresponding dependent variables. They have to be treated as exogenous variables, too.

Figure 1: Postulated causal model of performance on volumetric analysis calculations. The residual random variables are denoted by U_2, U_3, U_4.



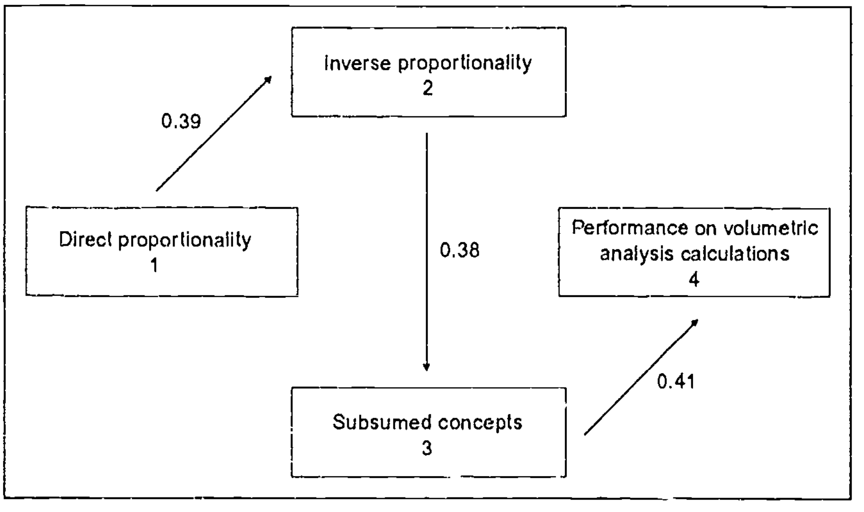
Especially for the group of students using algorithms without understanding, the path coefficients were estimated as summarized in the following table.

Table 1: Estimates of the path coefficients

Path coefficient	p_{43}	p_{42}	p_{32}	p_{31}	p_{21}
Estimate	0.42	–0.04	0.35	0.06	0.39

Although the applied statistical test indicated that a good model fit was achieved by the postulated path analytic model, further investigations of the individual path coefficients showed that the path connecting DP and SC ($\hat{p}_{31} = 0.06$) as well as the path connecting IP and PVAC ($\hat{p}_{42} = -0.04$) might not be essential for explaining the test scores for the nsidered group of students. These paths were then deleted from the hypothetical model and the path coefficients were again estimated based on this reduced model. This so-called trimmed model with the estimated path coefficients is represented in Figure 2. It also provided a plausible representation of the collected data.

Figure 2: Trimmed model with estimated path coefficients



Summarizing the results obtained for the group of chemistry students using algorithms without understanding, the authors stated that the performance of these students on volumetric analysis calculations is not directly influenced by proportional reasoning strategies.

An analogous path analytic approach indicated for the second group of students, however, that their performance on volumetric analysis calculations is influenced by proportional reasoning strategies.

The authors used the program LISREL by Joreskog and Sorbom (1978) to conduct the path analyses. Programs for carrying out a path analysis are also implemented for instance in SYSTAT. But each program package can be utilised which offers the possibility to conduct a regression analysis or to calculate empirical correlation coefficients.

3. Introduction to meta-analysis

In this chapter, an introduction to the principles of meta-analysis is given where at first the basic idea of this method is described. In addition, the advantages and the limitations of meta-analyses are discussed before dealing with special methods relevant in our context.

Concerning the realization of a meta-analysis using a program package, it should be mentioned that no special procedures for conducting meta-analyses are implemented in

ERIC

standard statistical packages such as SAS, BMDP, or SPSS. But in many cases a meta-analysis can be handled similarly to the analysis of an ordinary data set. It must only be taken into account that the 'data' used for a meta-analysis are already statistical measures. Thus, investigators with experience in using statistical program packages should also be able to carry out meta-analyses by using the same package. Most of the computations which are necessary to calculate estimators of a common correlation presented below can be carried out on a pocket calculator, although the author must admit, that such calculations could be burdensome for a given practical problem.

3.1 The principles of meta-analysis

Typically, a certain research problem is investigated in several empirical studies which are conducted independently by different investigators. Under such circumstances, a procedure for combining the results of the related studies is of interest to increase the evidence in them, because a single empirical study does not in general yield a solution of a major problem. One possibility to cumulate the findings of different studies consists in just a narrative discussion of these results. Another and obviously a more effective way of combining evidence across studies is a statistical reanalysis of the statistical results gained from different individual studies. A procedure which combines results from different studies dealing with a related or even the same research question in such a way is called a meta-analysis. This name was introduced by Glass (1976), although the first papers dealing with the combination of statistical significance tests and with the combination of estimates of treatment effects date back to the 1930s (c. f. Tippett, 1931; Fisher, 1932; Cochran, 1937).

This type of reanalysis of statistical results gained in other investigations is presented in our context of 'Correlation and Causality', because it can help to increase the evidence in observed correlations by aggregating correlation coefficients across studies. It cannot, however, confirm a causal-effect-relationship, but it is possible to support a hypothesis concerning a certain relationship if this association is detected with a similar intensity in different independent studies.

At a first glance, meta-analysis seems to be a simple and effective tool which can always be used to increase the statistical evidence. It has, therefore, gained importance in disciplines such as educational, psychological, or social sciences. But there is a number of crucial points which has to be taken into account when conducting a meta-analysis (c. f. Hedges, Olkin, 1985; Hunter, Schmidt, 1990). An overview of such aspects can be found in Schmid, Koch, and LaVange (1991). This review paper is essentially referred to in the following discussion.

The authors state that a meta-analysis includes all statistical advantages of a study with a large sample size. This implies the chance of getting more precise estimates of the effect investigated in each study. In addition, it is helpful when investigations of certain subgroups are of interest. Usually, such a procedure cannot be recommended because of the small sample sizes available in each subgroup e. g. caused by gender with respect to the special effect. Combining several studies can therefore facilitate a subgroup analysis. Furthermore, it can occur, that an effect is not clearly observable in the individual studies, although it exists in fact. Here, a meta-analysis can also help to detect this effect because of the increased sample size. Finally, a meta-analysis allows a critical compari-

ERIC

son and judgement of different studies which yields additional information on the underlying common research problem.

However, the meta-analyst has to ensure that the studies are in fact comparable, that means, he or she has to ensure that the studies are indeed dealing with the same or at least a similar research question. Moreover, the effect of interest should be measured using the same scale in each of the studies to be combined. Having no common scale complicates the method of combining the evidence across different studies, which is especially the problem in educational, psychological, or social research as pointed out e. g. by Hedges and Olkin (1985). This is in the very nature of variables of interest in such fields of research. Here, the chosen scale to measure variables such as the pupils' attitude toward school depends heavily on the responsible investigator, because such variables, having a psychological structure, lack a natural scale of measurement.

Besides the fact, that the studies to be combined should be very similar, Schmid, Koch, and LaVange (1991) also recommend to include only studies with a high quality in a meta-analysis although this point gives rise to controversial discussions in the literature. The authors conclude, that "in any case, the criteria for including studies in the meta-analysis should be well defined before the literature search is begun".

But a meta-analyst should be aware of an additional problem when searching for studies to be reanalyzed and this concerns the so-called publication or availability bias. Obviously, a meta-analysis can only include studies which are published or in some other way available. Thus, the question arises how much the outcome of a meta-analysis is biased by not considering the results of studies which are not published. This aspect is also discussed by the above authors. Especially, it is pointed out that in general only studies with significant results are submitted and published. Thus, it is nearly impossible to learn about all those studies with null results concerning the research question of interest, but especially these studies could change the outcome of a meta-analysis. This argument is perhaps not of importance when dealing with a meta-analysis of studies in educational research, because it can be seen from publications in this field of research that negative findings are very often reported (see also Hunter, Schmidt, 1990, pp. 506 – 514).

Since a meta-analysis is only a reanalysis of statistical results derived by other investigators, it lacks control of the study design and of the data. It can also hardly be judged if the statistical methods applied in each study are used in a proper way. Hence, the statistical measures to be combined could be based on inappropriate methods and consequently they could be invalid. Involving such invalid results in a meta-analysis can yield a respectable bias in the overall measures.

Further points to be considered in a meta-analysis also concern the more statistical aspects of such a procedure. They are extensively discussed for instance in Hedges and Olkin (1985) as well as in Hunter and Schmidt (1990). A number of important aspects can also be found in the review paper of Schmid, Koch, and LaVange (1991). These points are summarized below.

Since the overall measures are often weighted averages of the estimates gained in each individual study, it is necessary to decide how the weights should be chosen. On the one hand, one could argue that each available study should be included in the meta-analysis with the same weight. This implies that each estimate e. g. has the same statistical quality independent from the quality and the sample size of each individual study. On

ERIC

the other hand, the weights can be adjusted with regard to the size of each study or to its relative quality. Whereas it is difficult to reflect the individual quality of each study in a numerical weight, a sample size adjustment is easy to handle. Usually, the large studies get a larger weight, because it is assumed, that the results of a large study are more precise in a statistical sense. These proposals for a weighting procedure are only rough hints, more sophisticated methods can be found in the literature for the different statistical measures to be combined across studies.

Many statistical procedures are based among others on the assumption of independence of the individuals being considered. This assumption carries over to the individual studies when combining the estimates of effect in a meta-analysis. Although it is often simply taken for granted that this assumption is fulfilled, the different studies in a special field of research are typically based on the same background as well as on results of former studies and are therefore often intercorrelated. To take this problem into account, Schmid, Koch, and Lavange advise to examine the data for time, center, and investigator effects before conducting an analysis which needs the assumption of independence.

The last point to be mentioned here deals with the heterogeneity of the results gained from studies to be combined. Following the arguments of Schmid, Koch, and Lavange, heterogeneity does not constitute a major problem if it is just the outcome of a random process or only caused by a different scale used to measure the variables of interest supposed the scale can be retransformed. But if the heterogeneity cannot be explained, it is difficult to interpret or even senseless to estimate, for instance, a common effect, because in such a situation the studies can perhaps not reasonably be described as sharing a common effect. If this is the case, it may be unavoidable to abandon the assumption of one single parameter. Then, it is often more informative not to combine the results by aggregating, but to discuss the reasons for heterogeneity. Is there, for instance, a varying composition of the population under consideration or are there influential variables which are not taken into account in each individual study? One possibility to check the assumption of homogeneity consists in using an appropriate statistical test before pooling e. g. different estimates. A variety of such tests for different statistical measures to be combined can be found e. g. in Hedges and Olkin (1985).

All these aspects mentioned above and certainly further points related to each individual problem should always be kept in mind when drawing conclusions from a meta-analysis.

3.2 Estimation of a common correlation

In the following, we focus our interest on the estimation of a common correlation which is assumed to underlie a series of k studies dealing with the same topic. Even though this is only one aspect being of interest in a meta-analysis of several studies, merely an overview of some approaches can be given in this paper. The cited literature contains a more extensive discussion of the methods presented here and should be referred to for further details on this topic.

Thus, we consider a series of k independent studies, in each of which the correlation $\varrho_{X_iY_i}$ between two continuous random variables (r.v.s) X_i and Y_i, $i = 1, ..., k$, is estimated. The correlation $\varrho_{X_iY_i}$ is given by

ERIC

$$\varrho_i := \varrho_{X_i Y_i} = \frac{\text{Cov}(X_i, Y_i)}{\sqrt{\text{Var}(X_i)}\sqrt{\text{Var}(Y_i)}}, \qquad i = 1, ..., k,$$

where Cov (X,Y) denotes the covariance of two r.v.s X and Y and Var (X) the variance of a r.v. X. It is supposed that there exists a common underlying population correlation ϱ, i. e.

$$\varrho_1 = \varrho_2 = ... = \varrho_k = \varrho.$$

There are different possibilities to derive an estimator of ϱ. The simplest method of estimating ϱ consists in calculating the arithmetic mean of the individually estimated correlation coefficients in each study. But if the studies vary in size or seem to be of different precision, some kind of weighted means appears to be more appropriate. Thus, most of the approaches are based on weighted linear combinations of the independent sample correlation coefficients r_i or of transformations of these estimators.

3.2.1 Combined estimators based on Bravais–Pearson correlation coefficients

Let us denote the j–th pair of observations in the i–th study as x_{ij} and y_{ij}, $i = 1, ..., k$, $j = 1, ..., n_i$, then the Bravais–Pearson correlation coefficient $r_{X_i,Y_i} =: r_i$ for the i–th study is defined as

$$r_i := \frac{\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.})}{\sqrt{\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i.})^2 \sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2}}$$

with

$$\bar{x}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij} \quad \text{and} \quad \bar{y}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}, \quad i = 1, ..., k.$$

The empirical correlation coefficient indicates the strength of a linear relationship between two variables. In general, an estimator of ϱ can be obtained as

$$\hat{\varrho} = \sum_{i=1}^{k} w_i r_i \quad \text{with} \quad 0 \le w_i \le 1 \quad \text{and} \quad \sum_{i=1}^{k} w_i = 1,$$

where the weights can be chosen

(a) as $w_{i1} = 1/k$, then each study gets the same weight and $\hat{\varrho}_1$ is just the arithmetic mean:

$$\hat{\varrho}_1 = \sum_{i=1}^{k} r_i / k,$$

(b) as

$$w_{i2} = \frac{n_i}{\sum_{j=1}^{k} n_j},$$

then each study is weighted with its sample size relative to the total sample size, i. e. a larger study gets a larger weight and $\hat{\varrho}_2$ is given as

ERIC

$$\hat{\varrho}_2 = \frac{\sum_{i=1}^{k} n_i r_i}{\sum_{j=1}^{k} n_j}, \text{ or}$$

(c) as

$$w_{i3} = \frac{\dfrac{n_i}{(1-\varrho_i^2)^2}}{\sum_{j=1}^{k} n_j/(1-\varrho_j^2)^2}, \qquad i = 1, \ldots, k,$$

i. e. each r_i is weighted with the inverse of its asymptotic variance, which can be derived under the additional assumption of a bivariate normal distribution of X_i and Y_i for each i, i = 1, ..., k.

Under the above assumptions, the last weights yield the best combined estimator of ϱ based on $r_1, \ldots, r_k$ with regard to the variance of the combined estimator of ϱ, but these weights are unknown, since they depend on the true values of ϱ_i, i = 1, ..., k. If, however, the assumption of homogeneity is indeed fulfilled, the weights w_{i3} reduce to w_{i2} for all i = 1, ..., k. Nevertheless, Hedges and Olkin (1985, p. 230) recommend not to use a linear combination of $r_1, \ldots, r_k$, unless the sample sizes in each study are extremely large.

3.2.2 Combined estimators based on z–transforms

Instead of combining the individually estimated correlations coefficients r_i, a combination of Fisher's z–transforms of r_i is usually used to estimate ϱ. The z–transform of r_i is given as

$$z_i := z(r_i) := \frac{1}{2}\log\frac{1+r_i}{1-r_i}, \qquad i = 1, \ldots, k.$$

It has the advantage that its asymptotic variance does not depend on the underlying unknown parameter ϱ_i. Thus, we can calculate a combined estimator $\hat{z}$ based on z_i with weights

$$w_{i4} = \frac{(n_i - 3)}{\sum_{j=1}^{k}(n_j - 3)},$$

where $1/(n_i - 3)$ is the asymptotic variance of z_i, i = 1, ..., k. The estimator $\hat{z}$ can then be retransformed to get an estimator of ϱ denoted by $\hat{\varrho}_4$:

$$\hat{\varrho}_4 = \frac{(e^{2\hat{z}} - 1)}{(e^{2\hat{z}} + 1)}.$$

Another possibility for choosing the weights is based on power considerations of a statistical test for the test problem $H_0: \varrho \le 0$ versus $H_1: \varrho > 0$, where the test statistic is given by

$$\sum_{i=1}^{k} w_i z_i .$$

ERIC

It can be shown, that the weights w_{i5} with

$$w_{i5} := \frac{(n_i - 3)\,(\zeta + \varrho/2\,(n_i - 1))}{\sum_{j=1}^{k} (n_j - 3)\,(\zeta + \varrho/2\,(n_j - 1))}, \quad i = 1, \ldots, k, \quad \text{with} \quad \zeta := \frac{1}{2} \log \frac{1+\varrho}{1-\varrho}$$

yield the greatest power at a given alternative $\varrho > 0$ (Viana, 1980).

Since these weights again depend on the unknown parameter ϱ, Viana (1980) suggests an approximation of w_{i5} for $-0.7 < \varrho < 0.7$:

$$w_{i6} := \frac{(n_i - 3)\,(1 + 1/2\,(n_i - 1))}{\sum_{j=1}^{k} (n_j - 3)\,(1 + 1/2\,(n_j - 1))}, \quad i = 1, \ldots, k,$$

which is independent of ϱ. The estimators $\tilde{z}$ and $\hat{\varrho}_6$ are then defined as

$$\tilde{z} = \sum_{i=1}^{k} w_{i6}\, z_i \quad \text{and} \quad \hat{\varrho}_6 = \frac{(e^{2\tilde{z}} - 1)}{(e^{2\tilde{z}} + 1)}$$

3.2.3 Combined estimators based on unbiased estimators

An alternative to the z–transforms z_i is given by the unbiased estimators of ϱ_i, which only exist in form of an infinite series (Olkin, Pratt, 1958). Thus, we take the following approximations $G(r_i)$ instead of the exact unbiased estimators (c. f. Hedges, Olkin, 1985, p. 225)

$$G(r_i) := \frac{r_i + r_i\,(1 - r_i^2)}{2\,(n_i - 3)}, \quad i = 1, \ldots, k,$$

to derive a combined estimator of ϱ:

$$\hat{\varrho}_7 = \sum_{i=1}^{k} w_{i7}\, G(r_i) \quad \text{with} \quad w_{i7} = \frac{(n_i - 1)}{\sum_{j=1}^{k} (n_j - 1)}$$

(Viana, 1980).

3.2.4 The maximum likelihood estimator

The last possibility to obtain an estimator of ϱ to be mentioned here is based on the maximum likelihood method. The maximum likelihood estimator $\hat{\varrho}_{ML}$ of ϱ can be obtained numerically as the solution of the following equation (Hedges, Olkin, 1985, p. 234)

$$\sum_{i=1}^{k} \frac{n_i\,(r_i - \hat{\varrho}_{ML})}{1 - r_i\,\hat{\varrho}_{ML}} = 0\,.$$

3.3 A test of homogeneity

In addition to the above presented methods for deriving an estimator of the common correlation ρ, it is often useful to check the assumption of homogeneity by a statistical test. Again let us denote the sample correlations for each study as r_i and their z–transforms as z_i, $i = 1, ..., k$. Then, we get a test procedure based on the test statistic $T(z_1, ..., z_k)$ with

$$T(z_1, ..., z_k) = \sum_{i=1}^{k} (n_i - 3)(z_i - \hat{z})^2, \qquad \hat{z} = \sum_{i=1}^{k} w_{i4} z_i$$

in the following way: for a given significance level α, $0 < \alpha < 1$, reject the hypothesis of homogeneity if $T(z_1, ..., z_k)$ exceeds the $100 \cdot (1 - \alpha)$ percent point of the chi–square distribution with $k - 1$ degrees of freedom.

But even, if the hypothesis is not rejected, the statistical test cannot be interpreted in such a way, that the hypothesis is true. i.e. that the correlations are indeed homogeneous. Such a result should only be used as an indicator, that the assumption of homogeneity might be true. For a discussion of this test see also Hedges and Olkin (1985, p. 235).

3.4 An example

Let us consider the following example which discusses the relationship between teacher indirectness and student achievement. This research question is part of several studies on teaching effectiveness which have been conducted by Flanders (for a detailed description see Flanders, 1970). The following table contains some short information on the five studies which are of interest here.

Table 2: Information on five studies on teaching effectiveness (Flanders, 1970, p. 390)

Number	Year data collected	Location	Number of teachers	Grade level and subject	Outcome variable
1	1959 – 60	Minnesota	16	8th grade, 1 hour mathematics	attitude and achievement
2	1959 – 60	Minnesota	15	7th grade, 2 hour English–social studies core	attitude and achievement
3	1964 – 65	Michigan	30	6th grade, self–contained	attitude and achievement
4	1965 – 66	Michigan	16	4th grade, self–contained	attitude and achievement
5	1966 – 67	Michigan	16	2nd grade, self–contained	attitude and achievement

Among other variables Flanders considered the teacher indirectness as possibly influential variable on the students' achievement, where the teachers' "...indirect influence consists of soliciting the opinions or ideas of the pupils, applying or enlarging on those opinions or ideas, praising or encouraging the participation of pupils, or clarifying and accepting their feelings" (Flanders, 1967, p. 109). Table 3 on the next page summarizes the

ERIC

observed correlations between these two variables for the studies described in Table 2.

Table 3: Bravais–Pearson correlation coefficients between an observational measure of teacher indirectness and adjusted student achievement (Flanders, 1970, p. 394)

Study number	1	2	3	4	5
Grade level	8th	7th	6th	4th	2nd
Correlation coefficient r_i, $i = 1,\ldots,5$	0.428	0.481	0.224	0.308	-0.073

In the following, a meta–analysis of these five studies will be conducted to estimate a common correlation based on the observed correlations given in Table 3. Before estimating a common correlation, the test statistic of the described test of homogeneity will be calculated. For this purpose, it is necessary to derive the z–transforms of the observed correlations r_i, $i = 1, \ldots, 5$. The z–transforms and additional measures which are necessary to calculate an estimator of the common correlation are given in Table 4.

Table 4: Measures for calculating estimators of the common correlation based on five studies

Study number	n_i	r_i	z_i	$G(r_i)$	w_{i1}	w_{i2}	w_{i4}	w_{i6}	w_{i7}
1	16	0.428	0.457	0.441	0.2	0.172	0.167	0.168	0.170
2	15	0.481	0.524	0.496	0.2	0.161	0.154	0.155	0.159
3	30	0.224	0.228	0.228	0.2	0.323	0.346	0.342	0.330
4	16	0.308	0.318	0.319	0.2	0.172	0.167	0.168	0.170
5	16	−0.073	−0.073	−0.076	0.2	0.172	0.167	0.168	0.170

The test statistic results in 2.838, which is less than $9.488 = \chi^2_{4,\,0.95}$, the 95 percent point of the chi–square distribution with four degrees of freedom. Although the assumption of homogeneity cannot be rejected by the used statistical test, the individually estimated correlation coefficients give rise to some further considerations. A comparison of r_5 with the four other estimated correlations yields that its value of −0.073 differs essentially from r_1, r_2, r_3, and r_4, which are all greater than 0.2. This difference can perhaps be explained by the different grade level of the pupils in this study compared to the other studies where the pupils are at least at the fourth grade while this 'outlying' correlation is observed for the second grade. If indeed this study does not fit intrinsically with the other four studies, it should be excluded from the meta–analysis. For illustrative rea-

Table 5: Measures for calculating estimators of the common correlation based on four studies

Study number	n_i	r_i	z_i	$G(r_i)$	w_{i1}	w_{i2}	w_{i4}	w_{i6}	w_{i7}
1	16	0.428	0.457	0.441	0.25	0.208	0.200	0.201	0.205
2	15	0.481	0.524	0.496	0.25	0.195	0.185	0.186	0.192
3	30	0.224	0.228	0.228	0.25	0.390	0.415	0.411	0.397
4	16	0.308	0.318	0.319	0.25	0.208	0.200	0.201	0.205

ERIC

sons, the combined estimators of ϱ are, however, calculated and tabulated in Table 6 for both cases.

For the following calculations the fifth study is now omitted. In Table 5, the necessary measures to calculate the test of homogeneity and estimators of ϱ are given.

The test of homogeneity conducted for the remaining four studies yields

$$T(z_1, \ldots, z_4) = 1.248 < 7.815 = \chi^2_{3,\,0.95}\,.$$

Again, the assumption of homogeneity cannot be rejected.

Table 6: Estimators of a common correlation based on six different approaches for five (situation A) and four studies (situation B)

Estimator	Situation A	Situation B
$\hat{\varrho}_1$	0.274	0.360
$\hat{\varrho}_2$	0.264	0.334
$\hat{\varrho}_4$	0.277	0.347
$\hat{\varrho}_6$	0.277	0.347
$\hat{\varrho}_7$	0.271	0.342
$\hat{\varrho}_{ML}$	0.273	0.338

In general, only slight variations can be observed for the estimates given in Table 6 which are derived using the different approaches presented above. Especially, the two estimators based on the z–transforms yield nearly the same estimate of ϱ.

Comparing the estimates in situation A with those in situation B, it can be seen, that the estimates in A are all less than those in B, which is not unexpected since the fifth study is excluded in situation B and especially this study is connected with a negative observed correlation. Another consequence of the exclusion of the fifth study is that the estimates in B correspond better with the individually estimated correlations $r_1, \ldots, r_4$.

If we restrict our interest on the estimates in situation B, it should be mentioned that all estimates are lying between 0.334 and 0.347 except the unweighted arithmetic mean $\hat{\varrho}_1$ with a value of 0.360. Thus, it could perhaps be stated that $\hat{\varrho}_1$ seems to slightly overestimate the true correlation for the given data.

Since this example only serves as illustration, a more extensive discussion does not seem to be necessary.

4. Discussion

As discussed in the preceding chapters, there is a common problem underlying investigations of assumed causal relationships among variables being involved in the research question of interest. This problem is related to the fact that observed correlations do not in general imply a causal connection between the considered pair of variables. That means, a causal ordering cannot be developed only based on the estimated correlation coefficients or on the partial correlations. Especially, partial correlation coefficients can yield misleading results (see e. g. Duncan, 1970; 1975, pp. 22 – 24). But, as it is pointed out by Duncan (1975, Chapt. 2), correlations can be used when reasoning in the other direction, i. e. when e. g. rejecting a causal structure because the observed correlations do not fulfill the conditions of the hypothetical model. Notice again, that, however, it cannot be concluded only from observed correlations coming close to the conditions of the postulated causal scheme, that this scheme is true since there will always be alternative models which are also compatible with the collected data.

In summary, a path analysis and a meta–analysis are two different approaches which can be both helpful when investigating linear relationships among several variables. While a path analytic model and the involved statistical methods can be used to substantiate or to reject a certain causal structure, a meta–analysis of independent correlation coefficients cannot verify a supposed causal relationship, but it can increase the evidence in it.

All statistical methods presented above are based on certain assumptions which could not be discussed extensively in this paper. In a given practical situation, it is absolutely necessary to check if the underlying assumptions are sufficiently fulfilled before a special method is applied because a inappropriate application of statistical methods can yield invalid results. Of course, no decisions and no conclusions should be based on such misleading results.

References

Anamuah–Mensah, J., Erickson, G. & Gaskell, J. (1987). *Development and validation of a path analytic model of students' performance in chemistry*. Journal of Research in Science Teaching, vol. 24, 723 – 738.

Blalock, H. M. Jr. (Ed.) (1971). *Causal models in the social sciences*. Chicago: Aldine–Atherton.

Cochran, W. G. (1937). *Problems arising in the analysis of a series of similar experiments*. Journal of the Royal Statistical Society, Supp., vol. 4, 102 – 118.

Duncan, O. D. (1970). *Partials, partitions, and paths*. In: E. F. Borgatta (Ed.), Sociological methodology 1970. San–Francisco: Jossey–Bass.

Duncan, O. D. (1975). *Introduction to structural equation models*. New York: Academic Press.

Fisher, R. A. (1932). *Statistical methods for research workers*, 4th ed. London: Oliver & Boyd.

Flanders, N. A. (1967). *Teacher influence in the classroom*. In: E. J. Amidon and J. B. Hough (Eds.), Interaction Analysis. Theory, research and application. Reading: Addison–Wesley.

Flanders, N. A. (1970). *Analyzing teaching behaviour*. Reading: Addison–Wesley.

Glass, G. V. (1976). *Primary, secondary, and meta–analysis of research*. Educational Researcher, vol. 5, 3 – 8.

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta–analysis*. San Diego: Academic Press.

Heise, D. R. (1969). *Problems in path analysis and causal inferences*. In: E. F. Borgatta (Ed.), Sociological methodology 1969. San Francisco: Jossey–Bass.

Hunter, J. E. & Schmidt, F. L. (1990). *Methods of meta–analysis. Correcting error and bias in* *research findings*. Newbury Park: Sage.

Joreskog, K. G. & Sorbom, D. (1978). *LISREL IV: Analysis of linear structural relationships by the method maximum likelihood – user's guide*. Chicago: National Educational Resources.

Kang, K. M. & Seneta, E. (1980). *Path analysis: an exposition*. In: P.R. Krishnaiah (Ed.), Developments in Statistics. New York: Academic Press.

Land, K. C. (1969). *Principles of path analysis*. In: E. F. Borgatta (Ed.), Sociological methodology 1969. San Francisco: Jossey–Bass.

Li, C. C. (1977). *Path analysis: A primer*, 2nd ed. Pacific Grove: Boxwood Press.

Olkin, I. & Pratt, J. W. (1958). *Unbiased estimation of certain correlation coefficients*. Annals of Mathematical Statistics, vol. 29, 201 – 211.

Schmid, J. E., Koch, G. G. & LaVange, L. M. (1991). *An overview of statistical issues and methods of meta–analysis*. Journal of Biopharmaceutical Statistics, vol. 1, 103 – 120.

Tippett, L. H. C. (1931). *The method of statistics*. London: Williams & Norgate.

Viana, M. A. G. (1980). *Statistical methods for summarizing independent correlational results*. Journal of Educational Statistics, vol. 5, 83 – 104.

Wright, S. (1921). *Correlation and causation*. Journal of Agricultural Research, vol. 20, 557 – 585.

Wright, S. (1934). *The method of path coefficients*. Annals of Mathematical Statistics, vol. 5, 161 – 215.

SUMMARY OF THE PLENARY DISCUSSION

Holger Eybe, Universität Dortmund, Germany

First of all, the studies Iris Pigeot used as examples were not chosen from an educationalist point of view, they were merely regarded as good examples with reasonable results for illustrating the application of the presented statistical methods. Iris Pigeot's contribution gives the statistical background for Michael Piburn's study. Thus, educational issues are not mentioned in this text.

It was discussed how it can be concluded from the estimated path coefficients whether the postulated model provides a plausible representation of the data. Three methods were mentioned. First, researchers who often use path analytic models will get experienced in judging the values of the estimated path coefficients. Second, it is possible to estimate the path coefficients of the residuals and compare them to the path coefficients of the potentially influential variables. If the estimates of the residuals are too large then a relevant variable might be missing in the model. Third, statistical tests can be used to check the models. However, the problem of multiple testing might occur if there is quite a number of models to be tested. Therefore, it was suggested only to check a model if other considerations, e. g. practical experience, indicate that the model is a good representation of the data.

It was pointed out that conducting a path analysis is a stepwise procedure. First, the variables have to be identified and arranged in a path diagram. The next step is the data collection. Then, it has to be checked whether the model is compatible with the data. If this is not the case, researchers should always be ready to dismiss their path diagram and construct a new model. A trimmed model can be used if some of the variables the model contains are not important for explaining the outcome variable. A new path analytic model is not required then. However, it is also possible that the data do not fit in with the model because relevant variables are missing. These variables could be found by discussing the model with other researchers and statisticians. Also, residual variables could be identified then. Generally, the results of a study can be improved by consulting statisticians or other researchers.

The qualitative and the quantitative approach should be combined. Tests with individual students can be regarded as substantial work to find hypotheses. Statistical tests and computer programs should sometimes be avoided as they may be not adequate. For example, studies with a large number of variables should not be path analysed. It is a better method to conduct a descriptive study then.

The connecting paths in the trimmed-model example were determined by the results and backgrounds of other studies. It would be possible to change the direction of the 'arrows', but the emerging estimates would be different. Also, the model would not be chronological in terms of the problem-solving process anymore.

As for the reliability of the results of this example, the numbers given by the authors were used without discussing their adequacy from a statistical point of view. Authors often do not give any information on the validity of their results. A discussion with statisticians could show to what extent the statistical results can be interpreted or generalised.

"WORK" AND "HEAT" IN TEACHING THERMODYNAMICS

Peter van Roon, Universiteit te Utrecht, The Netherlands

1. Introduction

"Work" and "heat" are words that sound familiar to everyone from common parlance. In ordinary life, they refer to distinct, only vaguely related issues. They are also important scientific concepts, particularly considering their mutual relationship as formulated by the First Law of Thermodynamics.

In science education there are difficult concepts. This is demonstrated, among other things, by a regular appearance of articles on problems in teaching and learning these concepts in secondary and tertiary education. For instance, articles by Zemansky (1970), Warren (1972), Heath (1974), Tripp (1976), Erickson (1979, 1980), Summers (1983), and Se–Yuen Mak and Young (1987) can be mentioned. Many of these teachers/investigators analyse the raised problems and offer some kind of solution.

Zemansky (1970) recognises "three main infelicities of expression: (1) referring to the 'heat in a body'; (2) using 'heat' as a verb; (3) combining heat and internal energy into one undefined concept 'thermal energy', which on one page means heat and on the next page means internal energy". Zemansky's solution consists of a thermodynamically "rigorous treatment of the First Law".

Tripp (1976) states that "the concept of heat is basic to an understanding of energy transfer and elementary thermodynamics. Unfortunately, many students approach the study of science with a complete misunderstanding of the concept of heat. This misunderstanding will be reinforced if any one of a large number of available general chemistry textbooks is consulted". He comes to the conclusion that "the most common misconceptions regarding heat relate to the idea that it is something which is a component of a system". Finally, he states that "the successful teaching of the concept of heat will be accomplished if a clear definition of the system, the surroundings and the boundary is presented followed by an emphasis on the necessity of a temperature *difference* between the system and the surroundings and the associated transfer of energy across this boundary".

Summers (1983) confirms that "a basic difficulty has always been that the term heat is often used in everyday language as though it signified something that a substance can contain. However, it is now widely acknowledged by both science teachers and textbook authors that this is to confuse heat with internal energy. The problem is undoubtedly linked to the use of the word heat as a noun, which in turn may have its roots in the early caloric theory in which heat was regarded as a substance". Summers suggests that "use of the word *heat* as a noun should be avoided". This is the very solution already rejected by Zemansky in 1970.

Se–Yuen Mak and Young (1987) reject Summers' solution: "one proposal is to avoid the word 'heat' as a noun. This proposal ... at first sight seems to solve the problem. Moreover it draws attention to 'heating' as a process rather than a state, v .ich is certainly a step in the right direction. At the very least it avoids referring to internal energy as heat. However, on close scrutiny, the proposal has a number of problems as well, not the least of which is the danger of falling into the opposite error of referring to heat as internal energy or transfer of internal energy".

With the exception of Erickson's study, all others mentioned before are not founded on empiric educational research. They are written by teachers of thermodynamics within a thermodynamic context, the very context into which they aim at introducing their students. However, it is questionable whether these students, not yet being accustomed to a thermodynamic context, are at all able to comprehend the solutions offered. It is even questionable whether they are able to comprehend the problems to which solutions are offered, since these problems are related to the very thermodynamic context.

This paper describes the first part of an empiric educational research into problems with teaching and learning of "work" and "heat" in a thermodynamic context, *taking the context (or contexts) of the students themselves as a starting-point.*

This research was 'generated' by a number of 'symptoms'. One of these symptoms was that freshmen during their introductory course in physical chemistry at the University of Utrecht were reported to have 'problems' with concepts of heat, work, internal energy and enthalpy. Another symptom was that students in a further stage of the chemistry curriculum, when asked to apply thermodynamic concepts to chemistry laboratory situations, were reported to be unable to do so.

I decided not to concentrate on efforts to remedy this symptoms, for instance by presenting mathematical heuristics for problem solving as described by Mettes and Pilot (1980), Mettes, Pilot, Roossink and Kramers-Pals (1980, 1981), and recently by Hamby (1990). Like Se-Yuen Mak and Young (1987) I do not want to treat thermodynamics as a set of related algebraic symbols and equations whilst ignoring the underlying thermodynamic concepts behind symbols and equations. Although I am not blind to the possibilities of such heuristics for giving insight in the 'organisation' of an existing thermodynamic context, I think that they are not very much of use to students in *developing* a thermodynamic context.

2. A conception of "context"

In my opinion, a process of teaching and learning is a *communication process* between a 'teacher' and one or more 'students'. Therefore, problems of teaching and learning can be conceived as *communication problems*. They can be seen as a result of the fact that 'teacher' and 'student' each speak from a different *context*, that they each speak a different language. The result is misunderstanding.

I hereby use the word context in a sense in which it is used in philosophy of language. Martin (1987) defines "context" as "a piece of language with a hole in it". In this research I use the following definition of "context of a concept":

- By "context of a concept X" I mean a coherent, lingual structure of related concepts (to which X itself belongs), constituting a certain (view on) 'reality' in a sense that each concept within this structure refers to, and gives meaning to, an element (entity) of this 'reality'.

Without a context, a word just stays a word; only in a context it can become a concept. So, in this research I look for context differences, causing communication problems between 'teacher' and 'student', resulting in problems of teaching and learning of introductory thermodynamic concepts.

The paper now continues with a description of "heat" and "work" in a thermodynamic context, not because I am directed towards *imposing* this context on the students but as a 'frame of reference' when identifying students' context(s). Ultimately, the *introduction* of the students into a thermodynamic context should in my opinion be the goal of any education in thermodynamics.

3. "Work" and "heat" in a context of introductory thermodynamics

The thermodynamic concepts *heat (q)* and *work (w)* are encountered, together with the concept *internal energy (U)*, as equivalent terms in a much used formulation of the *First Law*:

$$\Delta U = q + w$$

In this formulation other thermodynamic concepts implicitely play a role. The quantity ΔU for instance is the change of internal energy of a *thermodynamic system* going from one *equilibrium state* to another as a result of some interaction between the system and its *surroundings*. This 'interaction with change of internal energy' is usually denominated as *energy exchange* between system and surroundings. During this *process* some *state quantities*, characterizing the equilibrium state, will probably change their value too.

A *thermodynamic system* is conceived as an 'abstract description' of the material object one is interested in, for instance of a pump, an electric cell, a calorimeter, a reaction vessel, each with its contents. Within the framework of this 'description' the object concerned is first selected and separated in thought from the remainder of material reality and subsequently reduced to a set of relevant outward characteristics: the *state quantities*. Important ones are mass, pressure, temperature and volume.

In order to enable interaction between a thermodynamic system and the 'outside world' another system is selected. This second system is called *thermodynamic surroundings*. In what way system and surroundings are able to interact, is determined by the *boundary* between system and surroundings. The various devices, such as containers, pistons, membranes, partitions, which are used to enforce thermodynamic boundary conditions (to a certain degree) on the material objects 'described' as systems, are traditionally referred to as *walls*.

System and surroundings together form a 'super system', isolated in a sense that whatever happens within this 'super system' is assumed to have no influence on the rest of 'reality'.

A *coherent* set of values, substituted for state quantities and sufficient to specify the 'condition' of the system, is known as an *equilibrium state* of the system. Such a state is time–independent in a sense that all values of the state quantities are constant in time and there is no net material flux over the system boundary. *This time–independency is the main reason why the equilibrium state is the most important, perhaps even only important, state of classical thermodynamics.*

Two main state quantities of classical thermodynamics are *internal energy* and *entropy*. Internal energy is particularly important because it is *postulated to be a conserved quantity*. One of the formulations of the *First Law of Thermodynamics*, which actually should be conceived as a thermodynamic *principle*, states that the internal energy of an isolated system has a constant value. Now because the 'super system' of thermodynamic system and thermodynamic surroundings is isolated, as was stated before, this implies that the internal energy of the 'super system' has a constant value.

A change of the state of a system from a certain initial state to a certain final state is called a *process*. If a process is accompanied by a change in the internal energy of the system, this change is the result of some kind of interaction between the system and its surroundings. If this interaction is only the result of a temperature difference between system and surroundings it is denominated as *heat* (or, a bit sloppily, as *heat exchange*). All other, adiabatic, kinds of interaction are known as *work*. Usually, interactions between system and surroundings are combinations of heat and work. Since in classical thermodynamics changes of internal energy which are the result of radiation or nuclear reactions are not taken into consideration, heat and work can be seen as complimentary ways of interaction between system and surroundings; together they determine the change of the internal energy of the system.

So *heat and work denominate kinds of interaction between the system and its surroundings, accompanied by changes in the internal energy of the system.*

A certain process can often proceed in different ways. For instance, the process in which hydrogen and oxygen react to water can proceed both instantaneously (oxyhydrogen explosion) and controlled slowly (fuel cell). Such a way is called a *path*. There are usually many paths leading from a certain initial state to a given final state. For each of these paths the sum of heat and work remains the same, but the division over heat and work may be different. Therefore, although the *sum* of heat and work under all circumstances equals the change of the state quantity internal energy, both heat alone and work alone doesn't.

Finally I can say that while "heat" and "work" were words denominating important founding concepts of classical thermodynamics, *the thermodynamic quantities heat and work are no (changes of) state quantities. They are process quantities, meaningless in the one important thermodynamic state: the equilibrium state.* Perhaps this is one of the main sources of difficulties with these thermodynamic concepts.

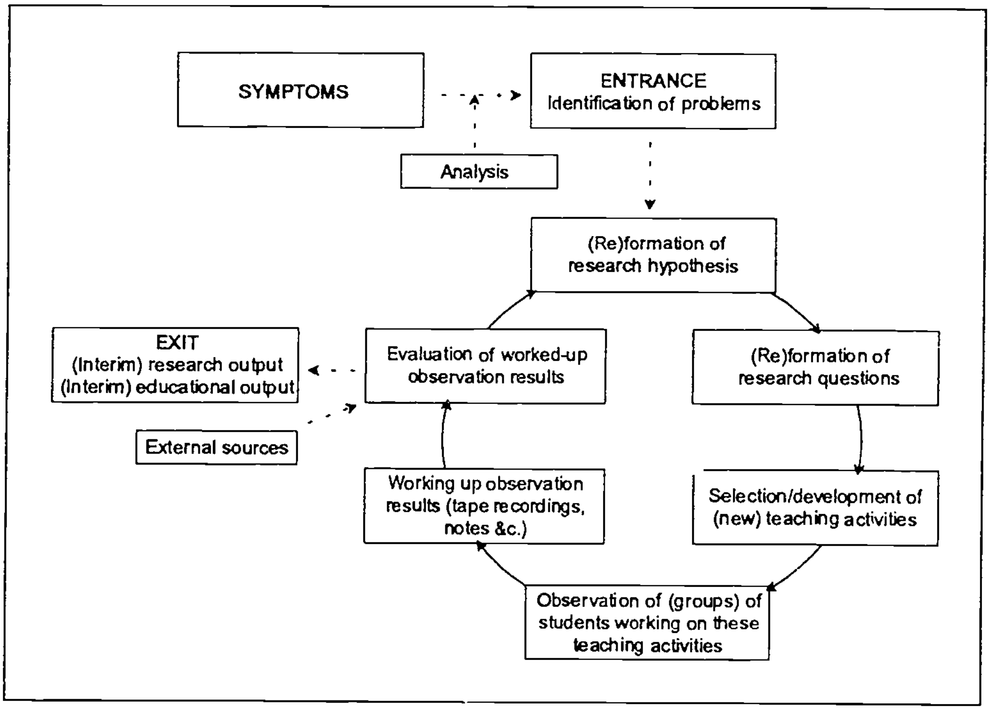
4. Educational environment of this research

The research described in this paper was carried out during a freshmen's introductory course in physical chemistry at the university of Utrecht. The course consisted of a series of twenty–eight two hours lectures on three different subjects: successively fourteen on chemical thermodynamics, seven on chemical kinetics and seven on electrochemistry, at a rate of two lectures a week. Each lecture was accompanied by a two hours tutorial session, in which students practised in smaller groups on problem solving, doing exercises on topics treated during the lectures. There were about six of these groups, each group consisting of about twenty students and one group teacher. This paper describes research performed during two successive introductory courses, viz. in 1986 and 1987.

It started in 1986 with the observation of only one group of students during all tutorial sessions on thermodynamics. It focussed on the acquisition of seemingly relevant students' enunciations on "heat" and "work", simply by listening to, and recording, the discussions of some small randomly composed subgroups of about four students, working on regular thermodynamic exercises. Since this research started in 1986 only a few weeks after the beginning of the tutorial sessions, and since especially during these first weeks exercises focussing on heat, work and the First Law were presented to the students, the research was continued in 1987. In this year tutorial sessions missed in 1986 were attended.

While working on their exercises, students were urgently asked to discuss between themselves the solutions they produced and the difficulties they experienced looking for these solutions. Subgroup discussions were recorded by means of a miniature tape recorder. Students were also asked to individually take notes and to hand over a copy of these notes to the 'observer'. The 'observer' too occasionally took notes of observations which might be relevant as background information. Tape recordings, student's notes and observer's notes together constituted the research material.

Figure 1: The cycle of research



This material was scrutinously analysed from the following "viewpoint": *"is it possible to point out any difference(s) between 'the' meaning of "heat" and "work" in a thermodynamic* *context and 'the' meaning (or perhaps more than one) given to these words by the observed students".* Observed differences of meaning were used as 'indicators' for context differences.

Of course, both 'the' thermodynamic meaning of "heat" and "work" and 'the' observed students' meaning, mentioned before, are *interpreted* meanings. However, these interpretations were intersubjectivated by way of triangulation (Berg, 1989; Maso & Smaling, 1990), both within my group and by 'external' experts on thermodynamics and/or education.

The empirical research, described in this paper, is *qualitative* research and, as is customary (Maso & Smaling, 1990) with this kind of research, is "meant to be cyclic, alternating the acquisition and interpretation of data in an interactive way" (fig. 1). Although the research was performed during two successive years, this whole research can, in my opinion, be qualified as introductory: as a 'cycle 0'. Main reason for this denomination is that during all two years this research was not yet based on one ore more explicitly stated research question(s).

5. "Work" and "heat" in collected students' enunciations

In this paper I can only present a very small sample of my research material to illustrate my results and conclusions. I selected two fragments from tape recordings that I regard as *'eloquent examples'* of 'the' meaning(s) students give to "heat" and "work" in their argumentations.

The recording fragments were written out into protocols and these protocols were subsequently translated from Dutch into English. In both protocols "Peter" is used for a non–student (mostly the group teacher concerned). All other names refer to individual students. But they are *fancy* names.

Protocol 1, heat reservoir conversation (1986)

John (J) and Peter (P) talk about the meaning of "heat reservoir". John just stated that: "I assumed that this is adiabatic .... ."

P *What is adiabatic?*

J *Ehm, this uh, uh now, uh here no heat, uh, net I mean, is absorbed.*

(silence)

P *Uh, you mean the heat reservoir and the system as a whole you look upon as a, an adiabatic ...*

J *No, only the heat reservoir.*

P *As a, an adiabatic system?*

J *Yes.*

P *I think that's strange because the heat reservoir is precisely for giving off heat or absorbing heat.*

ERIC

J *Yes, but I really mean that uh, it ... that it uh ... the heat absorbed equals the heat given off.*

P *That the heat given off by the system is absorbed by the heat reservoir and vice versa?*

J *Yes.*

P *Yes.*

J *Then you get that the ... q reversible uh, given off equals ... q reversible absorbed and together they are zero.*

(very long silence)

P *And that is both at the same temperature?*

J *Yes.*

P *Now, then in total, uh ... ,uh ... , now, but you said* ΔS *of the heat reservoir equals zero.*

(silence)

J *Uh, yes.*

P *That I do not understand ... because you say that the heat reservoir absorbs just as much heat as the system gives off or vice versa.*

J *No, I did say, uh, that the heat reservoir ... absorbs just as much heat as it gives off to the cylinder ...*

(very long silence)

J *or else, uh, the temperature doesn't remain constant too.*

P *OK, uh, now I really understand you. ... You actually say, you conceive the heat reservoir as a kind of buffer ... for heat.*

J *Yes.*

P *Now, it isn't, uh, meant that way and that has, uh, apparently been mentioned very little in the lectures and so, uh, I would skip that question. ... .*

In a thermodynamic context, a heat reservoir, originally introduced by Carnot to denote a very large 'container' of 'heat particles' (caloric), is a surroundings meant to by–pass one of the consequences of the law which was eventually known as the Zeroth Law of Thermodynamics This consequence states that two thermodynamic systems of originally different temperature, when brought into close thermal contact, in the end assume the same *intermediate* equilibrium temperature. However, a heat reservoir, conceived as being very much larger than the finite system with which it is in thermal contact, *imposes* its temperature on this finite system. It is supposed that this coupling does not disturb the internal equilibrium of the heat reservoir (Tisza, 1966). A thermostat can be conceived as a 'practical translation' of a thermodynamic heat reservoir.

John apparently conceives a heat reservoir as an 'adiabatic system' (*l.* 6 – 8) since, in his opinion, there is no *net* absorption of 'heat' by the heat reservoir (*l.* 2), and therefore no net 'heat exchange' between the heat reservoir and the remaining part of 'reality': "the heat reservoir ... absorbs just as much heat" (from elsewhere, for instance from an immersion heater) "as it gives of to the cylinder" (*l.* 27, 28) "or else the temperature doesn't remain constant" (*l.* 30). John speaks of a thermostat, *including* an immersion heater, as "a kind of buffer ... for heat" (*l.* 31 – 33). He refers to the heat reservoir with the word adiabatic. Apparently, "adiabatic" to him means "constant heat" in stead of "no 'heat exchange' possible". John appears to reason in terms of *conservation of heat*. To him "heat" seems to be a 'state quantity', "something in a body", and not a 'process quantity' (Zemansky, 1970).

While concentrating only on his heat reservoir, John doesn't seem to separate the *combination* of cylinder and heat reservoir from the rest of 'reality', as is necessary for them to become a combination of system and surroundings in a thermodynamic sense. For if the 'heat' transferred from the heat reservoir to the cylinder is compensated in his 'heat buffer', this 'compensating heat' must come from somewhere else outside his 'buffer' (*l.* 27,28). And this conflicts with the thermodynamic meaning of "surroundings".

Besides, John himself doesn't mention the words surroundings and system, not even when Peter does. He only speaks of heat reservoir and cylinder. Whether John's 'heat' refers to "heat particles" (caloric) or to "heat energy" remains unclear. Peter is unable to problematize "heat reservoir" with John; he just cuts off the discussion.

Protocol 2, work–to–heat conversation (1986)

Peter (P) teaches Anne (A) about a (thermodynamic?) relation between work, heat and energy.

A *After all, I must do work before uh ... before I can uh ... supply energy uh ... somewhere.*

P *In what sense do you use work?*

(very long silence)

A *Well before I can uh ... supply energy uh ... to, to something I must do work, it's simple as that.*

P *Eh, ... yes, I don't understand that ... uh ... I don't understand what ... what you want to tell me.*

A *You don't? Well, ... as far ... as far as I'm concerned that's not my fault, but uhh ... (grin) ... .*

A *Yes, I mean eh ...*

(very long silence)

A *here, if I want to give this thing (takes up an object) a certain en ... eh ... potential energy ...*

P *Yes.*

A *then I must do work by bringing it up like that (lifts the object over his head).*

P *Yes, I understand that.*

A *Yes, well now, that's exactly the same, isn't it?*

P *No ...*

A *Just now ... just now I said if I something ... want to supply energy to something, I must do work on it ...*

P *no!!!*

A *and now I say the same thing and now you do understand me.*

P *Yes, because that is a purely mechanical uh ... example ... and there you use ...*

A *O, yes, if I'm going to supply heat uh ...*

P *there you use the words potential energy and work*

A *Yes.*

P *and uh, I ... I know from mechanics ... Yes, ehm ... ehm ... the ... the work done is the difference in potential energy, yes, OK. But this is a uh ... thermodynamic problem! This concerns the relation ... between ... work, heat and energy, and here energy has the meaning of ... an ... energy state function! (silence) And that is not the energy in your ... mech ... mech ... in your sen ... mechanical sense.*

A *Yes but I say ... here I think then that that ... that I have to ... change that state function then because the degree of dissociation is zero at first ... so I only have* N_2O_4 *... and thus then I have to uh ... uh ... a part one sixth there of that* N_2O_4 *I have to transform into* $2\,NO_2$.

P *Yes.*

A *So, I have to supply that energy ...*

(silence)

A *and I can only do so by by ... somehow or other uh ... doing work so uh ... so that heat arises so that this ... this energy can be absorbed or something like that on such a way, do I know how ... else.*

P *Oh! ... Oh, so ... uh*

(silence)

P *if you want to uh ... uh, in an electricity ... yes I, I ... I am just imagining something. If you want to burn coal uh ... in an electricity generating station to ... uh ... drive the generator so you can generate an electric tension ... you first have to do some work on that coal ... to ... uh?*

A *Yes, but here it doesn't go all by itself.*

Anne has problems with a distinction between *mechanical energy* (kinetic and potential energy of a macroscopic body, for instance the 'system' as a whole) and *thermodynamic internal energy* (energy conceived as a state function) (*l.* 1, 2; *l.* 14 – 24, "the same" in *l.* 19 and *l.* 24 referring to *l.* 1, 2).

Peter tries to instruct her, from his thermodynamic context, where she is wrong and why she is wrong (*l.* 29 – 33) but apparently with no success (*l.* 41 – 43). Even using an ordinary life example of an electricity generating station doesn't avail (*l.* 46 – 49): Peter asks Anne whether she means that to obtain 'heat' from burning coal first some work must be done on that coal. He probably assumes this example to be so absurd as to convince Anne of her mistake. Much to his embarrassment however Anne agrees (*l.* 49, 50).

Anne reasons in terms of *conversion* in stead of conservation of energy: "I can only do so by somehow or other *doing work* so that *heat* arises so that *this energy* (*in casu* heat) can be absorbed or something like that ... ." (*l.* 41 – 43). She seems to imply that "*doing work*" ('work energy') generates (is converted to) "heat" ('heat energy'), relating "heat" to "energy" (heat is *a form of energy*). Speaking like that, she apparently conceives "heat" as a 'state quantity' and not as a 'process quantity'. This *work–to–heat conversion*, by the way, is familiar from expressions like "heat of friction".

ERIC

Peter and Anne seem to speak different languages. The result is misunderstanding. Peter's teaching is a complete failure.

6. Characterizing and naming students' context(s)

In a preliminary conclusion I stated that the observed students use "heat" as a 'state quantity' and not as a 'process quantity'. There are at least two scientific conceptions in history in which heat was used as a 'state quantity'. In the first one "heat" was conceived as *material*: "heat particles" (named "caloric" by Lavoisier in 1787 (Roller, 1950; Tisza, 1966)). I refer to this 'material heat' concept as *caloric heat concept.* In the second one "heat" was conceived as *energetic*: "heat energy" or "thermal energy". This conception was historically an immediate successor of the caloric conception, and a predecessor of classical thermodynamics (or perhaps even a first stage of classical thermodynamics). It also belonged to calorimetric thermochemistry (Mach, 1986). I refer to this concept as *energetic heat concept.*

In both conceptions I can distinguish a principle of *conservation of heat*. However, this principle is not identical for both. With regard to 'caloric' it may actually be conceived as an example of the principle of *conservation of matter*: 'caloric' heat *particles* are conceived to be indestructible. With regard to 'heat energy' it may be conceived as an example of the principle of *conservation of energy.*

"Work" can be used in a similar way both as a *mechanical work concept* of Newtonian mechanics and as a *thermodynamic work concept.* In both cases "work" is defined as "the vector product of some 'force' and some 'displacement' of that 'force'". One difference between these conceptions is the fact that a mechanical concept of work is generally used for 'motion' of a 'system' (macroscopic body) as a whole, whereas a thermodynamic concept is generally used for 'motion' of a wall or a part of a system (for instance of a piston in a cylinder, or of electric charge). Both conceptions of work are related to some principle of *conservation* of energy. However, only in a mechanical conception also an energy *conversion* principle (for instance potential energy to kinetic energy and *vice versa*) is used. Such an energy conversion principle is superfluous in classical thermodynamics since in there *only one* energy concept, "thermodynamic internal energy", is used. In non–classical, irreversible thermodynamics, by the way, mechanical work is simply included in thermodynamic work.

As an important aspect in the characterization of students' context(s) I see the relation between "heat" and "work" which they use. Although there exist *parallels* between a caloric heat concept and a mechanical work concept (conservation of caloric heat parallels conservation of mechanical energy; conversion of latent caloric to free caloric parallels conversion of potential energy to kinetic energy), both concepts are principally *separate* and *unrelated*. They have no 'common factor' which relates caloric "heat" to mechanical "work". There is no common context. Even where it is a matter of caloric heat generating mechanical work (for instance in Carnot's heat engines) it is the *motion* of the heat *particles* that generates the work. The heat particles themselves remain unchanged during this process, since caloric is conceived to be indestructible. There is *no heat–to–work conversion.*

This situation is different for an energetic heat concept. In this case, besides two sepa-

ERIC

rate conservation principles (conservation of heat and conservation of mechanical energy), there also is a 'common factor' which relates this heat concept to a mechanical work concept: "energy". "Heat energy" and "work energy" are conceived as different *forms* of energy. And these different forms of energy can be *converted* into one another. Energetic "heat" ("heat energy") and mechanical "work" are therefore related in one common context. I already situated conceptions like "heat of friction" in such a context. Since, as I mentioned before, this common context is also the context of calorimetric thermochemistry, an immediate predecessor of classical thermodynamics, I name this proto–thermodynamic context *"thermochemical context"*.

A thermodynamic context is *in itself* a common context of "heat" and "work". It has no need of 'forms of energy' which can be converted into one another. The internal energy of classical thermodynamics has no 'forms'. Internal energy is conserved, but not converted.

Although it is difficult to differentiate "work" as a thermodynamic concept and "work" as a mechanical concept, I conclude from my research material that the observed students still use "work" as a (kind of) mechanical concept. This conclusion is based mainly on the fact that, in my impression, these students relate "work" to "heat" by means of energy *conversion*. Protocol 2 offers but one example to found this impression.

I also conclude that the observed students' conception of heat is an energetic heat concept. I base this conclusion on the fact that I found both "conservation of heat" (in protocol 1) and "heat energy" (definitely in protocol 2 but probably also in protocol 1).

- ❑ I see so many parallels between students' context(s) and a scientific context which I named "thermochemical context" above that I characterize students' context(s) as thermochemical.

In my introduction I stated as my aim "teaching and learning of 'work' and 'heat' in a thermodynamic context, taking the context or contexts of the students themselves as a starting–point". Now this "starting–point" has been identified as a thermochemical context.

References

Berg, B. L. (1989). *Qualitative research methods for the social sciences*. Boston, Allyn & Bacon.

Erickson, G. L. (1979). *Children's conceptions of heat and temperature*. Science Education, vol. 63, 221 – 230.

Erickson, G. L. (1980). *Children's viewpoints of heat: a second look*. Science Education, vol. 64, 323 – 336.

Harnby, M. (1990). *Understanding the language: problem solving and the first law of thermodynamics*. Journal of Chemical Education, vol. 67, 923 – 924.

Heath, N. E. (1974). *Heating*. Physics Education, vol. 9, 490 – 491.

Mach, E. (1986). *Principles of the Theory of Heat*. Dordrecht, D. Reidel Publishing Company.

Martin, R. M. (1987). *The Meaning of Language*. Cambridge, Mass., MIT Press.

Maso, I. & Smaling, A. (1990). *Objectiviteit in kwalitatief onderzoek*. Meppel, Boom.

Mettes, C. T. C. W. & Pilot, A. (1980). *Over het leren oplossen van natuurwetenschappelijke problemen*, Doctoral Thesis. Enschede, University of Twente.

Mettes, C. T. C. W., Pilot, A., Roossink, H. J. & Kramers-Pals, H. (1980). *Teaching and learning problem solving in science, part I: a general strategy*. Journal of Chemical Education, vol. 57, 882 - 885.

Mettes, C. T. C. W., Pilot, A., Roossink, H. J. & Kramers-Pals, H. (1981). *Teaching and learning problem solving in science, part II: leaning problem solving in a thermodynamic course*. Journal of Chemical Education, vol. 58, 51 - 55.

Roller, D. E. (1950). *The Early Development of the Concepts of Temperature and Heat*. Cambridge, Mass., Harvard University Press.

Se-Yuen Mak and Young, K. (1987). *Misconceptions in the teaching of heat*. The School Science Review, vol. 68, 464 - 470.

Summers, M. K. (1983). *Teaching heat-an analysis of misconceptions*. The School Science Review, vol. 64, 670 - 676.

Tisza, L. (1966). *Generalized Thermodynamics*. Cambridge, Mass., MIT Press.

Tripp, T. B. (1976). *The definition of heat*. Journal of Chemical Education, vol. 53, 782 - 784.

Warren, J. W. (1972). *The teaching of the concept of heat*. Physics Education, vol. 7, 41 - 44.

Zemansky, M. W. (1970). *The use and misuse of the word "heat" in physics teaching*. The Physics Teacher, vol. 8, 295 - 300.

COMMENTARY OF THE DISCUSSANT

Gabriela Jonas, Universität Hamburg, Germany

The subject physics is one of the least liked subjects of our students. Within the subject physics again thermodynamics is one of the areas our students do not like and they are not interested in. Furthermore, thermodynamics is very hard to understand and not closely related to our everyday life experience. So I am always amazed when someone chooses thermodynamics for research. In your paper, Peter van Roon, you explain the process of misunderstanding as a communication problem. I certainly agree with your point of view and I also agree that the professor should be the one who has to change his language. I wonder, why you have chosen the chapter thermodynamics to approach the communication problem between students and the professor? You apply the communication problem for university students and their professors. Does it also apply for teachers and students at school? How do you evaluate your result: "... the students have a thermochemical point of view"? Is that what you would have expected? Are you satisfied with it or would you wish to have got a different result? If so, do you already have an idea what should be changed in teaching thermodynamics at school level? That is where students got their point of view from.

To overcome the language problem between students and professors, do you already have an idea how thermodynamics aspects could/should be taught at university level? Which words should the professor use? I cannot really see a way yet because we still need to teach the content of thermodynamics; which other words could be taken? Your research so far has been — in my opinion — a very important startpoint for more, very much needed, research in the process of teaching and learning thermodynamics. I have only one more question regarding to your study group: How representative is your group of students? They are all students of chemistry. Can it be that their concepts are a bit different than concepts of students who are not studying science?

I like your investigation, first to look at students concepts and then to focus on the teacher. I think your ongoing research on that topic can help to improve the teaching and learning process of thermodynamics and so I hope thermodynamics will be liked more by teachers as well as by students.

SUMMARY OF THE PLENARY DISCUSSION

Petra Beuker, Universität Dortmund, Germany

In Peter van Roon's opinion thermodynamics is rather easy because it is a closed object. Thermodynamics should not be taught in secondary schools where the students are introduced to e. g. ΔH. This conception hinders them when they go to university. Entropy and enthalpy only make sense in the thermodynamical context. It is too abstract for students of 14/15 years and therefore it might be enough to tell them about the heat of reactions. Thermodynamics should only be taught at university.

Peter van Roon was not able to compare his results to other results already published in the literature, because there are not any yet. He was the first to write an article about this subject and hopes that it will be accepted. As for the validity he said that students are all individnals and that the same things never happen twice.

There is a communication problem between teachers and students at university and school level. It is impossible to use other words instead of the special words e. g. in thermodynamics. Therefore it is easier to teach a subject that is not known at all. When students come to university they already have a pre–conceived view of 'heat and work'. The question is how they can get the same definition as the teachers. Peter van Roon said that their conceptions cannot be changed and that pre–conceptions are not necessarily misconceptions. The teacher can only try to change their point of view by asking the students about their ideas. Education has to be changed according to the cyclic system. Peter van Roon's study is the entrance part of the cycle (see hand–out). In his opinion there should be an agreement between the teacher and the students. A reference to this could be e. g. that the teacher steps back in his language to the students' level. Otherwise there will be misunderstandings and no progress can be made. Until now the teachers do not have any information on how to change their language and to synchronise.

Peter van Roon said that this is one of the basic questions in his work. He co–operates very closely with a professor in physics who also teaches thermodynamics. His aim is to show him that his work does not lead to the understanding of the topic, but only to passing the exams. Another problem is that there is no sound theoretical base in thermodynamics which are in fact thermostatics. Therefore the starting situation for finding a better language is not so good.

One of the participants said that there are problems and obstacles concerning a common

ERIC

languge of students and teachers (at uiversity), determined by the everyday colloquial experince of the former group:

(1) The name "thermodynamics" is already misleading, it stems from Carnot's paper "Sur la puissance motrice du feu" (about the moving power of fire). The study of equilibrium states which do **not** change in time gives an **upper** bound to the real efficiency. The Carnot **idealisation** corresponds to quasistatic processes (slow changes of equilibrium states) and thus has **zero** power output, whereas a **real** machine is characterised by a kind of **maximum** power principle.

(2) An ideal thermal reservoir has to consist of infinitely many degrees of freedom to allow heat exchange without temperature change.

(3) A strictly isolated system with finitely many degrees of freedom will not show thermodynamic behaviour (e. g. approach to equilibrium). Its entropy is temporally constant, both in the classical axe (Lionville's Theorem in phase space) and also quantum–mechanically (Unitary invariance of the trace).

METHODOLOGICAL ASPECTS OF UNIVERSITY OF DORTMUND STUDIES OF STUDENTS' CONCEPTIONS IN CHEMISTRY

Elisabeth Schach, Universität Dortmund, Germany

1. Introduction

Planned research studies have several components that follow each other in a logical order. The steps involved in an empirical study are shown in Table 1. The chronology of steps of a research project usually implies the repetition (Table 1, steps marked by vertical bars) of some of these steps as the research question evolves.

Table 1: Components of empirical studies in chemistry education

- Research question (specific wording of major and related question)
- Related studies (aims, methods, findings)
- Methods (type of study, sample, instruments, field phase, data)
- Pilot phase (small field phase to evaluate the studies feasibility and the potential relevance of the findings)
- Field phase (small field phase to evaluate the studies feasibility and the potential relevance of the findings)
- Field phase (study data collection phase)
- Estimation, comparisons, relationships
- Results

consistent with — not consistent with

(reasons) — prior knowledge — (reasons)

further studies

- Generalisability } Conditions
- Repetition of study } Conditions
- Results

consistent with — not consistent with

(reasons) — prior knowledge — (reasons)

further studies

Study methods may be successively improved by preceding each new study by feasibility and pilot phases. Analyses of methodological aspects of each study further improve the understanding of study methods and thus, help to augment the design of the following one. Routine analyses of methodological aspects include the calculation of rates of missing data for each variable, the ascertainment of inconsistencies across variables and the analyses of missing values by placement of questions in the instrument. Studies among classes of students are associated with the problem of clustering of observations, an aspect which requires monitoring with respect its effect on study results. Furthermore, investigations among volunteers, as for example high school students, should be examined with respect to the effect this might have on the conclusions from such studies.

2. Two types of studies

Focussing on methodology, two basic types of studies may be useful in chemistry education. One type attempts to draw conclusions from random samples of students in order to generalise the results to all students of a particular kind.

The second type of studies focus is on comparing two or more groups of students, e.g. with respect to selected performance characteristics. Students in the respective comparison groups need not be drawn at random from the student population. However, student assignment to groups should be on the basis of randomisation.

3. Methods of quantitative studies

Let us focus on *population studies* first. These studies are performed in order to obtain population estimates. Such estimates might be aimed at ascertaining:

- the proportion of high school students with misconceptions in chemistry?

  or

- the proportion of advanced chemistry students with a specific type of misconception?

As already mentioned, the study design for such an investigation calls for a random sample of students of a region or a school system.

Following the estimation of rates from such random samples, such rates are valid estimates of the respective population statistics. Biases may be introduced by non response of students or teachers, who had been drawn into the sample but for whom data could not be collected.

Comparative studies among student groups are designed in order to compare two or more groups cross-sectionally or at two time points. Research questions might be:

- do high school students in elementary chemistry courses use different problem solution strategies than advanced chemistry students?

  or

- do advanced chemistry students solve specific problems with less misconceptions than when they were beginners?

Design methods for comparative studies differ from those employed in population studies. The design calls for obtaining homogeneous group(s) of students in order to compare them with respect to the so-called treatment. Such a treatment might be the administration of a test.

As it is difficult to assign equivalent students to each one of the treatment groups out of the pool of students that are available for the study, it is generally recommended that students be assigned to treatment groups at random. This procedure is used in order to minimise biases that might otherwise be the consequence of systematic assignment. The validity of study results hinges on the degree of achievement of group equivalence by random assignment of students to the groups. If the equivalence of study groups is achieved, then observed differences between treatment and control groups are attributable to the effect of the treatment. When the study groups resemble groups of the general population, then the results derived from such a study may be generalised to a broader subgroup of the population as compared with the case when study groups are composed of select subjects with rare characteristics.

4. Characteristics of studies among high school students

As compared to the prerequisites of population or comparative studies, the reality of studying high school students' performance is usually not in agreement with these principles. Studies are frequently carried out by volunteer teachers and/or volunteer students. The effect of this violation of prerequisites is that estimates of population rates/percentages will be biased, as Schach (1992) showed for estimates of smokers in the general population.

Identical tests are often administered to all students of a class and observations of all pupils of a class are subsequently entered into analyses. When students of one class are more homogeneous than students of different classes with respect to performance and possibly selected personal characteristics, these homogeneities are usually not taken into consideration in analyses.

Due to these limitations, estimation of rates can not be justified for these types of studies because the observations were not obtained for a random sample of students. Comparative analyses are appropriate, if the comparison groups are uniform with respect to all characteristics except for the specific 'treatment'. Such a treatment may consist of a test.

While the fact that education research in Germany will have to be content with co-operating teachers and volunteer students (instead of random samples of each), the assignment of students to test problems is an aspect of study design that may be controlled by investigators.

Analysis aspects are further facets of research methodology that may be controlied by investigators. Even though it might be difficult to explain to participating classes, that only a few students are selected for a research study, analyses of random students from all co-operating classes should be performed routinely in order to assess the impact of using all students or random samples of them on estimates of relationships. When all students of a class are used, relationships (such as correlation coefficients) are overestimated. Therefore. investigators might wish to learn about the order of magnitude of this bias (see Figures 1 and 2 for specific estimates).

5. Balancing the assigning test sets to students (of a class)

When designing a field study one wishes to obtain as much information as possible while using up a minimum of resources. The question might be: is this aim realised better by assigning identical test sets to all students of a class or is the assignment of different test sets to a class to be preferred. The latter holds true, as the multitude of students' problem solution strategies is better represented by varying test sets within a class. Assigning the same set of test problems to all students of a class is inefficient, since homogeneity of solution strategies is expected to be greater within than across classes. Another obvious advantage of such assignments is that chances of cheating are reduced.

Let us assume that student assignment to test sets be varied systematically, as part of the study design strategy. Then, the following aspects might be required:

(1) balancing of the assignment of topics to positions in the test set in order to control for decreasing probabilities of correct problem solution when moving problems from front to back positions in the test set,

(2) balancing the occurrence of problems within topics

(3) avoiding to hand out equal test sets to neighbouring students of a class,

(4) providing for the use of incomplete clusters of test sets in the analyses.

An example is given in order to demonstrate an assignment that fulfils several of the above requirements.

Let there be three topics, 5 problems for each topic and two problems to be assigned to each student (of a particular study); topics are not repeated in one test set. Furthermore, let us assume that each teacher receives 4 test sets (cluster).

There are 6 permutations of three topics into sets of two problems each, namely:

12, 13, 23, 21, 31, 32

and a selection of random orders of these sets is:

23	12	21	31	32	13
31	21	23	12	32	13
32	12	13	21	31	23

Let us denote the problems for by

topic 1 by	1,1	1,2	1,3	1,4	1,5
topic 2 by	2,1	2,2	2,3	2,4	2,5
topic 3 by	3,1	3,2	3,3	3,4	3,5

and let us choose a random start for problems of each topic (underlined).

We then assign problems to topics according to random orders of tuples of topics in a circular fashion beginning with a randomly chosen problem. Then a few of the test sets are:

2,2 + 3,5	3,2 + 2,5	2,2 + 3,5
1,3 + 2,3	1,1 + 3,3	1,4 + 2,3
2,4 + 1,4	3,4 + 1,2	3,1 + 2,4
3,1 + 1,5	2,1 + 1,3	1,5 + 3,2

If we review the placement of topics by positions in the above test sets, we obtain:

Topic 1 in position 1 : 4 times
Topic 1 in position 2 : 4 times
Topic 2 in position 1 : 4 times
Topic 2 in position 2 : 4 times
Topic 3 in position 1 : 4 times
Topic 3 in position 2 : 4 times

The assignment shows that 9 problems occur twice and six once in 12 test sets. Thus, requirements (1) to (3) are fulfilled by the previous test set.

If only 3 and 2 and 4 test sets were returned out of each block of 4 test sets (which might be realistic since student numbers vary between classes), we obtain the following numbers of topics in positions 1 and 2 (9 test sets), respectively:

Topic 1 in position 1 : 4
Topic 1 in position 2 : 1
Topic 2 in position 1 : 3
Topic 2 in position 2 : 4
Topic 3 in position 1 : 2
Topic 3 in position 2 : 4,

i. e., despite of dropouts, we observe, each topic at least once in each position in the analysis set. This shows that random ordering of problems in sets of test problems sent to each teacher, reduces the impact of dropouts on returned test sets and thus on the dataset for analysis. Furthermore, requirement (4) is fulfilled by this procedure.

In addition to this positive effect on test sets that are going to be available for analyses, this procedure enables the investigators to fix the number of desired single test sets of a particular kind or to predetermine the number of combinations of sets that are desirable for specific types of analysis. The determination of the minimum number of sets or of combinations of sets is an optimisation problem, which may be solved mathematically. The minimum number of sets for a particular problem could be set on the basis of the accuracy requirements for a statistical estimate that shall be derived from the respective test problem.

Thus, planning of test sets offers the opportunity to introduce specific features into the study design. This opportunity is used too infrequently, in spite of the fact that balancing of characteristics or minimisation of observation numbers is possible.

The advantage of designed test sets is that stated study design objectives may be fulfilled. Furthermore, as student success rates decrease as positions of test problems increase within the set, this aspect of potential bias for study results is controlled by the design.

6. Effect of clustering of students in a class on success rates in classroom tests

When analysing teachers' contributions to study data sets, the choice is between using the data of all students of a participating class or just one random student's data of each class. The decision to analyse just one student from a class not only reduces the data set substantially, but also results in an increase in the variability of the observations. Both effects are undesirable. Figures 1 and 2 demonstrate these effects for 100 observations.

Figure 1: Success rates grade 10. All students and one student per class compared.

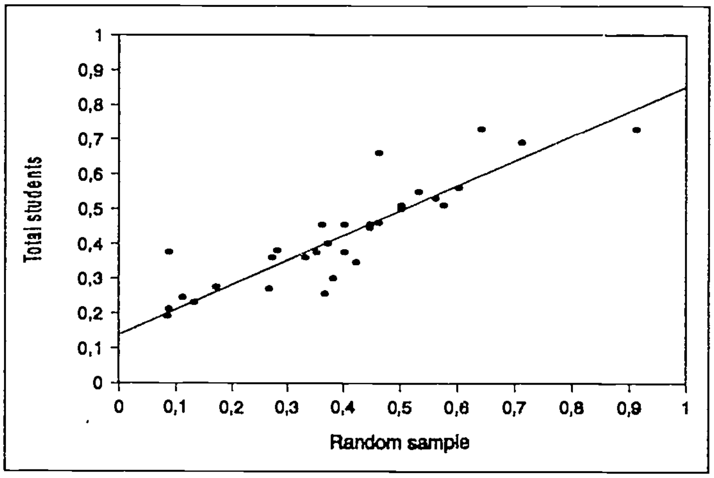


Figure 1 shows the relationship of success rates (percentage of problems answered correctly) for one randomly chosen student of a class and all students of a class (average success rate). While the success rates of individual students vary from less than 0.1 to > 0.9, the average success rates for complete classes range from 0.2 to 0.8. This latter reduction of variability results in shorter confidence bands and higher correlation coefficients, if the class–wise observations are used for such estimates, as compared with the respective statistics based on data of individual students.

Figure 2: Success rates grade 12/13 (advanced level). All students and one student per class compared.

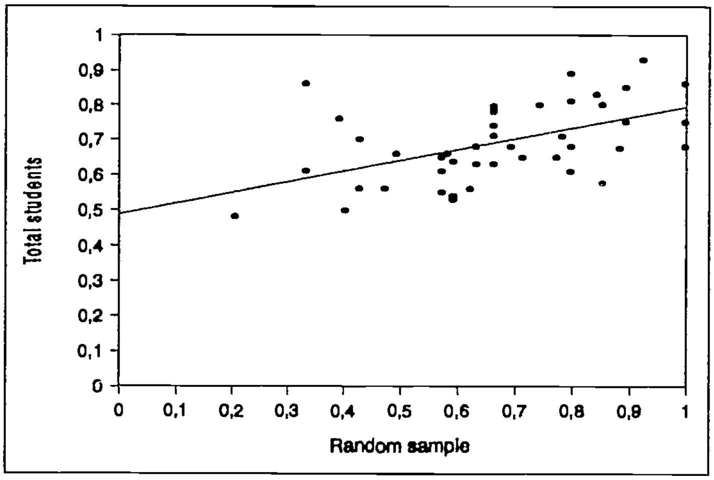


Figure 2 demonstrates the same phenomenon for success rates of advanced chemistry students. Success rates for randomly chosen students range form 0.2 to 1; total students' success rates vary much less, namely between 0.5 and 0.9. Thus, when average success rates of students per class are entered into analyses instead of observations on single students, the effect of clustering is even more severe for students of higher grades. This shows that the within–class homogeneity is larger for classes of advanced student (as compared with classes of elementary students). Thus, estimates of population standard deviations and correlation coefficients are biased, if they based on all observations of all classes.

7. Strategies of University of Dortmund studies of misconceptions in chemistry

Given the methodological limitations of studies among volunteers (students and teachers), study aims have to be in agreement with these limitations. While investigations aiming at estimates of rates (i.e. number of science students with misconceptions among the student population of a certain age) can not be estimated without bias on the basis of such data sets, comparisons of the performance of students may be carried out.

In summary the procedures that were used in empirical studies of student perceptions in chemistry (Schmidt, 1992), are characterised as follows:

- volunteer teachers and students as study participants and
- impracticality of administering tests to only one student in a class.

Given these limitations of the study design, we attempted to:

- avoid population estimates from study, rather performed group comparisons,
- ascertain estimates that compare subgroups of overall study group,
- systematically vary student assignments to topics and test problems in order to avoid identical test sets for student neighbours,
- place test topics in each position of the test in a balanced fashion (in order to control position as confounder of study results),
- routinely carry out analyses for key variables on the basis of all students <u>and</u> random samples of students of participating classes (in order to learn about attenuation effects on correlation coefficients).

Results of such studies may be generalised to the extent that this is justified on the basis of the study's methodological characteristics, such as subjects, instruments, and selection procedures.

References

Schach, S. (1992). Methodische Aspekte bei der telefonischen Bevölkerungsbefragung. Grundsätzliche Überlegungen und Ergebnisse einer empirischen Untersuchung. In: S. Schach und G. Trenkler (Hg.), Data Analysis and Statistical Inference. Festschrift in Honour of Friedhelm Eicker. Department of Statistics. University of Dortmund. Bergisch Gladbach: Verlag Josef Eul.

Schmidt, H.-J. (1992). *Harte Nüsse im Chemieunterricht*. Frankfurt/M.: Diesterweg.

A Case Study on Students' Difficulties Applying the Brønsted Theory

Hans–Jürgen Schmidt, Universität Dortmund, Germany

1. Introduction

For naming substances, reactions and concepts chemists usually choose labels that give information on the ideas that are behind these terms. The label *equilibrium* refers to a balance between forward and backward reaction. It is the situation in which the forward reaction yields exactly the amount of substance that is used up by the backward reaction. However, the label *equilibrium* can lead to the misconception that the concentrations of product(s) and reactant(s) are equal (Hackling and Garnett, 1985). The label *oxidation* was originally limited to reactions in which oxygen took part. Today, changes in the oxidation numbers of elements involved indicate a redox reaction. However, oxygen is not necessarily involved in redox reactions. The idea of oxidation has changed whereas the label remained. There is a danger that students use the concepts redox reaction as if it was an oxidation in its original meaning (Garnett, Garnett and Treagust, 1990). The label *neutralisation* indicates that equivalent amounts of acids and bases are used up. However, if, for example, acetic acid and sodium hydroxide react, the resulting solution is not neutral. The label *neutralisation* is often misleading and students, therefore, predict an incorrect result for this reaction (Schmidt, 1991).

2. Background

The Brønsted theory states that acids are proton donors. Bases are proton acceptors. Acids and bases always occur as a pair. The acid HA donates a proton and the base A^- remains. The base B^- accepts the proton and forms the acid HB:

(1) $HA \rightleftharpoons H^+ + A^-$

(2) $B^- + H^+ \rightleftharpoons HB$

The complete reaction is

(3) $HA + B^- \rightleftharpoons A^- + HB$

Thus, acids and bases do not destroy each other and the label neutralisation no longer has its original meaning.

Four acid–base pairs can be formed from the acids and bases in equation (3). The acid HA and the base A^-, as well as the acid HB and the base B^- are yoked together. They are called *conjugate* acid–base pairs. Some authors use the word *corresponding* instead of

ERIC

conjugate. The label corresponding focuses on another aspect. The acid–base pairs, HA / A^- and HB / B^-, are independent, but correspond with one another. They exchange a proton, like a message. The four acid–base pairs of equation (3) consist of two acid–base pairs that are conjugate and two that are not conjugate.

The present investigation was initiated by a pre–test conducted by the Examination Committee of the American Chemical Society. In a multiple–choice pre–test senior high school students were to identify a conjugate acid–base pair in the following equation:

(4) $NH_3\,(aq) + H_2O\,(l) \rightleftharpoons NH_4^+\,(aq) + OH^-\,(aq)$

The test item contained the following options:

- the correct answer (NH_4^+ / NH_3)
- the acid–base pairs that are not conjugate (H_2O / NH_3 and NH_4^+ / OH^-)
- the two bases OH^- / NH_3
- the two acids NH_4^+ / H_2O

The pre–test showed that students especially preferred *one* of the acid–base pairs which is not conjugate, namely NH_4^+ / OH^-. We were interested in the reasons for their choice. Therefore we assigned equation (4) and

(5) $H_3O^+\,(aq) + HCO_3^-\,(aq) \rightleftharpoons H_2O\,(l) + H_2CO_3\,(aq)$

(6) $NH_3\,(aq) + HCO_3^-\,(aq) \rightleftharpoons NH_4^+\,(aq) + CO_3^{2-}\,(aq)$

to senior high school students. They were asked to find the conjugate acid–base pairs for the reactions and to give reasons for their answer. The students' most frequent error in (4) and (5) was to consider the non–conjugate acid–base pairs NH_4^+ / OH^- and H_3O^+ / HCO_3^- as conjugate. The written comments indicated that students had chosen pairs of ions, i. e. particles with opposing electrical charges. In equation (6) the pair NH_4^+ / CO_3^{2-} consists of an ion with a single positive charge and an ion with a double negative charge. The incorrect answer NH_4^+ / CO_3^{2-} was second in frequency of choice. The most frequent incorrect answer was NH_4^+ / HCO_3^-. It could be seen from the comments that the students had tried to find a matched pair of ions, one with a single positive charge and one with a single negative charge. According to the result of this prestudy many students seem to consider a pair of ions with equal opposing electrical charges that somehow neutralise each other as conjugate acid–base pairs.

3. Problem

The aim of this investigation was to identify misconceptions senior high school students have about Brønsted's theory of acid–base pairs. It focused on whether students' most frequent incorrect answer is to regard pairs of ions with equal opposing electrical charges as conjugate acid–base pairs. Within the scope of the investigation multiple–choice questions with distractors that reflect students' misconceptions were to be developed.

4. Method

4.1 Instruments

Free–response questions and multiple–choice questions were used in this study. The questions contained the aforementioned equations (4), (5) and (6) plus the equations (7) and (8).

(7) $HCl\,(aq) + H_2O\,(l) \rightleftharpoons H_3O^+\,(aq) + Cl^-\,(aq)$

(8) $NH_3\,(aq) + HSO_4^-\,(aq) \rightleftharpoons NH_4^+\,(aq) + SO_4^{2-}\,(aq)$

The equations were included into two standardised texts. The following examples are representative of the two types of test items resulting from this procedure.

Test item 1

The equilibrium between ammonia molecules NH_3 and hydrogen sulphate ions HSO_4^- in an aqueous solution is described by the following equation:

$$NH_3\,(aq) + HSO_4^-\,(aq) \rightleftharpoons NH_4^+\,(aq) + SO_4^{2-}\,(aq)$$

Which particles form a conjugate acid–base pair, according to Br~nsted?

[A] NH_4^+ / NH_3 (conjugate acid–base pair)
[B] NH_4^+ / HSO_4^- (pair of ions)
[C] NH_4^+ / SO_4^{2-} (non–conjugate acid–base pair)
[D] HSO_4^- / NH_3 (non–conjugate acid–base pair)

Test item 2

The equilibrium between ammonia molecules NH_3 and hydrogen carbonate ions HCO_3^- in an aqueous solution is described by the following equation:

$$NH_3\,(aq) + HCO_3^-\,(aq) \rightleftharpoons NH_4^+\,(aq) + CO_3^{2-}\,(aq)$$

Which particle(s) is/are the conjugate base(s) to the acid NH_4^+ ?

[A] NH_3 as well as CO_3^{2-} (conjugate and non–conjugate acid–base pair)
[B] Only HCO_3^- (pair of ions)
[C] Only NH_3 (conjugate acid–base pair)
[D] Only CO_3^2 (non–conjugate acid–base pair)

The distractors were chosen according to the results of the pre–test. The remarks in brackets show how the distractors were selected and were not present on the test. Eleven test items of this type were used for the investigation.

4.2 Data collection and sample

The test was administered in the schoolyear 1991/92. Known teachers and teachers from school registers were asked to participate. 4,291 senior high school students (Gymnasium) completed the test. The response rate was 36 %. The students were asked to give reasons for their answers. In addition, a discussion with a group of 14 senior high school students on test item 1 was conducted. The students were on a day visit at our university and were not part of the test population. The talk was videotaped and the important parts were transcribed.

The curriculum for senior high school students is divided into elementary and advanced chemistry courses. These differ in the number of lessons per week (2 to 3 for elementary and 5 to 6 for advanced courses).

4.3 Design

The present study was part of a major investigation in which 122 questions were used. The questions were divided into 6 groups of approximately 20 items each. Every student received a test package of 6 test items which were laser-printed. Each test package contained only one item from each group. Each item appeared with the same frequency in each position of the test package. In order to achieve this distribution the following method was applied: First, the sequence of the 6 groups within the test package was randomly determined. In this procedure for each package every group could be drawn only once. Next, one test item for each of the six positions was randomly selected from each group.

4.4 Findings

The results of the study will be illustrated using the multiple choice questions 1 and 2. In Table 1 the distribution of students' answers among the options are presented. The incorrect answers that occurred were those that had been expected: the students considered a pair of oppositely charged ions and the two non-conjugate acid-base-pairs as the conjugate acid-base-pairs.

Table 1: Distribution of stud..nts' answers among the options: 12th and 13th class, elementary courses (e); 12th and 13th class, advanced courses (a); multiple choice item (m); text as ir question 1; text as in question 2; * = Conjugate and non-conjugate acid-base pair

Item	course	Options chosen, in %					
		correct answer	ion pair	non-conjugate	non-conjugate	no answer	number of students
1	e	40	17	18	10	15	82
	a	64	14	7	5	10	84
2	e	46	21	9	10*	15	68
	a	63	8	13	8*	10	80

The same incorrect answers were given in free-response questions as well as in multiple-choice questions. The reasons for this choice could be extracted from students' written comments. Here is a typical example of how students gave reason for the correct answer of item 1: *"NH_3 is a base because it accepts a proton. HSO_4^- is an acid because it donates a proton. Consequently, NH_4^+ is an acid ... and SO_4^{2-} is ... a base. NH_3 / NH_4^+ as well as HSO_4^- / SO_4^{2-} are corresponding acid–base pairs."* The following student had an extraordinary idea: *"Conjugation is the declension of verbs, that means you put one and the same verb into a different form. Therefore the conjugation of chemical substances would always have to involve the same atoms. Conjugation means, I suppose, ... the transformation of a base into an acid. Therefore the last three answers cannot be correct."*

The usual reasoning in favour of distractor B in item 1 was as follows: *"These two ions complement one another. The positive charge of the NH_4^+ ion is needed by the HSO_4^- ion, because a negative charge is missing. Thus HSO_4^- is the acid ion and NH_4^+ is the base ion."*

Similar comments could be found for distractor B in item 2: *"NH_4^+ has a single positive charge. Only HCO_3^- has a single negative charge."*

The following comment shows why in item 1 students opted for distractor C, a non-conjugate acid–base pair. *"... because ... HSO_4^- donated a proton, that is H^+, ... which was accepted ... by NH_3 ..., turning it into NH_4^+."* The next comment on the free-response version of item 1 shows the reasons why students were distracted by the other nonconjugate acid–base pair. *"NH_3 / HSO_4^- because this is a proton transfer from hydrogen sulphate to ammonia."*

The students who were on a day visit at our university first solved items 1 and 2 and then discussed their results. At the beginning of the discussion there was a vote on item 1. Student #1 voted for B and said: *"An NH_4^+ particle is a Brønsted acid and HSO_4^- is a base. The base on the left side of the equilibrium is connected to the conjugate acid on the right side."* In the subsequent discussion the students came to the end that A is the correct answer. Eventually they considered what thoughts could have led students #1 to the incorrect answer B. Student #2 remarked: *"...you take HSO_4^- and on the other side you see NH_4^+. Then you remember that ... acids and bases have something to do with protons. NH_4^+ has protons in excess, HSO_4^- has electrons in excess, that means it lacks protons. So you think that NH_4^+ functions as an acid and HSO_4^- as a base and that this is the conjugate acid–base pair. HSO_4^- and NH_4^+ seem to belong together, as if they somehow neutralised each other."* Student #1 replied: *"That's how I proceeded, ... I have seen the plus and the minus ..., but I have not considered the reaction equation"*. Student #3: *"This is what always happens when definitions are extended. Acid–base reactions used to be defined as reactions between H^+ and OH^-, ... a neutralisation. If this is extended to donator–acceptor reactions ... you may stick to the pattern of the plus and minus neutralisation."*

5. Discussion

Students seem to have two misconceptions about acid–base pairs. First, they confuse non-conjugate and conjugate acid–base pairs. Perhaps, they have not considered the difference between the two. They chose that acid–base pair that can be found on the same side of the equation as the conjugate pair. The most important misconception is to regard positively and negatively charged ions as conjugate acid–base pairs. The first step in this direction is to confuse electrons and protons. One student wrote: *"... NH_4^+*

can function as an acid because of its electron in excess." This may have been a slip of the tongue. In the following comment both terms are used: "*Bases are electron acceptors, acids are proton donors. H_2O is the conjugate base to OH^-.*" The second part of the first sentence could have led the student to the correct solution, however, he only referred to the first part. A student's comment that has already been mentioned in the results section was: "*HSO_4^- is the acid ion and NH_4^+ the base ion.*" Another student whose comment has also been mentioned above regarded, vice versa, NH_4^+ as Brønsted acid and HSO_4^- as Brønsted base. These students do not seem to have a stable conception about which particles to define as acids and which as bases. They simply concentrate on electrical charges as if they neutralised each other. Perhaps the label acid–base pair tempts students to think of neutralisation.

Each test item was assigned to about 200 senior high school students. As the items were randomly distributed to 4,253 high school students the results cannot be generalised beyond this group. In the course of our investigations it was often discussed whether the data should be based upon random samples. This would lead to a defined population. As long as the situation can be expected to be the same a replication study using this population would give the same results. In the present study in which teachers volunteered the result may be biased by certain characteristics of the volunteers. However, there is good reason to continue working with volunteers:

- It seems to be impossible to draw a random sample. If, for example, only 50 % of the teachers were selected at random respond one does not arrive at a random sample.
- Teachers who have already taken part in similar investigations are more likely to cooperate. They should have the better chance to influence their students to explain their reasoning strategies resulting in lower costs for the study. As our investigation involved the development of test questions, they can be used to validate the results with other populations.
- The study uncovered misconceptions, originating from chemistry itself. The hypothesis is that they can appear everywhere.

6. Conclusions

This study investigated the reasons why students chose the wrong answers to Brønsted acid–base tests. Two factors were identified. Students regarded non–conjugate acid-base pairs or positively and negatively charged ions as conjugate acid–base pairs. They confuse non–conjugate and conjugate acid–base pairs. The result suggests that chemistry teaching should address the following problems:

- In Brønsted's acid–base reactions four different acid–base pairs are involved. Only two of them are conjugate acid–base pairs.
- There is a difference between proton–transfer and electron–transfer reactions.
- Pairs of ions with opposing electrical charges can never be conjugate acid–base pairs.

If teachers are aware of students' misconceptions described above they will be better able to remove them.

References

Garnett, P. J. Garnet & Treagust, D. F. (1990). Implications of research on students' understanding of electrochemistry for improving science curricula and classroom practice. International Journal of Science Education, vol. 12, 147 – 156

Hackling M. W. &. Garnett, P. J (1985). *Misconceptions of chemical equilibrium*. European Journal of Science Education, vol. 7, 205 – 214

Schmidt, H.–J. (1991). *A label as a hidden persuader: chemists' neutralization concept*. International Journal of Science Education, vol. 13, 459 – 471

Acknowledgements

Theo Ashford †, American Chemical Society / Examinations Committee, provided us with pretest results. Serpil Cankay was associated collecting and analysing the data. Statistical advice was given by Elisabeth Schach. The project was funded by the *Deutsche Forschungsgemeinschaft (Schm. 701/3 -2)*. I would like to acknowledge their support.

SUMMARY OF THE PLENARY DISCUSSION

Holger Eybe, Universität Dortmund, Germany

It was pointed out that in Hans–Jürgen Schmidt's investigation there was no comparison between the groups. There was a differentiation between elementary and advanced course students. This differentiation, however, was only used because there are these two levels at secondary level education. It was merely tried to determine what strategies students use and what misconceptions occur. It was found that although students are very different in their abilities and personalities the same misconceptions and problem solving strategies occur everywhere. The study did not aim at gaining any quantitative results. Also, there was no intention to generalise the results.

The hypotheses for these studies are derived from analyses of questions from examination boards in the Netherlands or the UK. These tests provide new ideas about what misconceptions are around. It is an aim of the investigations to describe these conceptions and to reveal reasons for them. Thus, the studies could be regarded as descriptive studies.

As for the conclusions and implications of the study, it could be seen that problems on acid and bases sometimes trigger the concept neutralisation. This concept, however, cannot be applied to Brønsted's idea of acids and bases. But the problem solving behaviour is still influenced by this concept. Thus, it might be helpful if teachers discussed this problem with their students.

A Study of the Relationship between Conceptual Knowledge and Problem–solving Proficiency

Amos A. M. Shaibu, Ahmadubello University, Nigeria

Abstract

The relationship between conceptual knowledge and problem–solving is articulated in a number of theories and research studies e. g. Gagné (1985), Gabel et al (1984), Stewart (1982). Both teachers and curriculum planners are concerned that students acquire the capability to solve problems efficiently. In this study, the relationship between the conceptual knowledge of pre–degree science students in Nigeria and their ability to use such knowledge to solve contextual problems was investigated.

The sample comprised 190 students drawn from 5 colleges across the country. The average chronological age of the sample was 18.5 years. They were well motivated and, as shown by their achievement test results, above average in ability.

Structured paper–and–pencil tests in mechanistic organic chemistry were used as instruments for collecting relevant data which were analysed using the "MINITAB" statistical package. The results of the study showed amongst other things, that :

(a) the students lacked functional understanding of the logical inter–linkages amongst the pool of chemical concepts that they have acquired.

(b) while the students possessed most of the requisite conceptual knowledge; they were unsuccessful in solving problems that required such knowledge as pre–requisites.

(c) there seems to be a very weak link between the students' possession of conceptual knowledge and their ability to use such knowledge to solve contextual problems. The variance in the latter accounted for by the former was found to be only 20 %.

(d) the assumption, often implicitly held that students would develop desired problem–solving capabilities if only they acquired relevant conceptual knowledge needs to be approached with caution, especially in its use as pedagogic basis for instruction in science education.

Some of the issues raised by these findings; and suggestions for improving quality of science instruction in the school are highlighted .

1. Introduction

The development in students of the ability to solve problems is a major goal in science education. Two types of knowledge have been identified to underline problem–solving proficiency. These is conceptual and procedural knowledge (Stewart, 1982, Greeno, 1978). The former refers to knowledge of concepts, laws and theories of a particular domain as generally presented to the student through some form of instruction. The latter refers to the knowledge of general heuristics that underlines the ability to utilise conceptual knowledge in solving problems. It is the acquisition of functional cognitive stratcgies or self-management procedures that pave the way for execution of a solution to a problem. Woods (1991) observed that problem-solving involves a mental and attitudinal processes connected by strategy. The process of developing strategies for solving problems requires the modification or a translation of acquired conceptual knowledge to fit the demands of the problem. This is made possible where the structure and hierarchy of the conceptual knowledge is both adequate and relevant.

Both teachers and curriculum planners are concerned that students acquire the capability to solve problems efficiently. However, there seems to be an implicit assumption, on their part, that this capability can be attained if students acquire the relevant conceptual knowledge. This assumption is highlighted by the findings of Perez and Terragrosa (1983) in which they reported that students simply do not learn to solve problems. Also, the conception and the procedure of administration of most routine school examinations in which students are usually given problems to solve as a method of assessing their knowledge-base appears to be a typical manifestation of this underlying assumption. Similarly, the choice of instructional methods which are often didactic, theoretical and teacher-directed (Ajeyalemi, 1983) is often guided by this assumption.

2. Background literature

Problem-solving is defined in different ways by different authors/scholars e. g. Gagné (1978), Ausubel (1968), Woods (1991), Hayes (1991). The one by Ashmore et al (1979) appears concise and comprehensive, and thus used as the working definition regarding what we are to understand by problem-solving in this paper. They define it as the end-product of application of knowledge and procedures to a problem situation. Problem-solving, therefore calls for the ability to bring relevant conceptual knowledge to bear on a given or perceived problem such that a reasonable solution is produced at the end.

The relationship between conceptual knowledge and problem-solving ability is articulated in a number of theories e. g. Gabel et al (1984), Gagné (1985). In fact Gagné considered learning as a composite structural hierarchy with problem-solving at the apex.

The theoretical frame-work for this study derives mostly from Ausubel's (1968) theory of meaningful learning hence problem-solving. In it, Ausubel distinguished between rote and meaningful learning even though he saw each of these to be at the respective ends of a learning continuum. One major highlight of Ausubel's theory is that where students do not possess requisite subsumers in their cognitive structure, they essentially resort to learning by rote. The structure and hierarchy of concepts (knowledge) acquired in that circumstance is most often unhelpful in problem-solving situations. On the other hand, if/when students possess requisite subsumers, then the chances of fruitful processing of new information is maximised, leading to re-structuring or modification of existing knowledge, ultimately resulting in meaningful accommodation of the new knowledge. This, often transforms into a coherent cognitive pattern that is functionally viable for problem–solving.

It is therefore the case that while all students learn in one sense or another, some instructional environments foster meaningful learning while other do not. In the latter case, the students resort to rote learning which has no conceptual anchorage, and hence unviable, in most part, as pre–requisites for problem–solving.

3. Purpose of the study

Many studies have been reported in literature concerning students' understanding (or lack of understanding) of science concepts, for example, Mitchell and Gunstone (1984), Butts and Smith (1987), Finley et al (1982), Garnett and Treagust (1992), Shaibu (1988). Most of the results showed areas in which students experience conceptual difficulties. However, fewer reports seem to be available in literature about the relationship between students' conceptual knowledge and their problem–solving capabilities. It seems to be the view of many teachers that problem–solving skills and indeed proficiency in problem–solving can be attained by students through a process of acquiring conceptual knowledge. Selection of teaching methods and planning of general instruction is often guided by this assumption.

This study is an attempt to identify the relationship between the conceptual knowledge of pre–degree science students in Nigeria and their ability to solve problems based on such knowledge as necessary pre–requisites. Specifically, the study sought to:

(a) find out if the students possessed the relevant conceptual (chemical) knowledge to which they were previously exposed through a process of planned instruction; and which were identified in the study as prerequisites for solving a set of given problems;

(b) find out, based on (a) above, if the students were successful in solving the problems;

(c) determine the relationship between the students' conceptual knowledge and problem–solving proficiency.

4. Method

4.1 The sample

The sample comprised a total of 190 final year students of the schools of Basic Studies in Nigeria, who offer chemistry as one of their 3 major subjects which are covered at the same level as the GCE A'level. Five of the schools, spread across the country were involved in the study. The sample was obtained through the process of random selection. This was with a view to obtaining an unbiased sample, which according to Robson (1983) affords each member of the population an equal chance of appearing in the sample.

ERIC

4.2 Instruments

The instruments employed for the investigation consisted of structured, pencil–and–paper tests in organic reactions. There were two of them tests A and B. Test A comprised 50 multiple–choice items, which were later reduced to 40 after pilot–testing. The items were derived and organised after a careful study of the students' syllabus. Also a survey of their past examinations scripts for 5 consecutive years was undertaken. The syllabus gave guidance in mapping out the knowledge area included in the test and also the difficulty–level incorporated. The survey of the past examination papers gave insight into the students' pattern of solving problems, thus helping as a guide in the selection of the multiple–choice options i.e. the keys and distractors. This procedure was useful in ensuring that students selected the right options because they had the underlying knowledge of the students.

Test B which comprised 5 free–response problems in its final version, was built from (structured on) test A, such that the students required knowledge of test A to solve the problems in test B successfully. It was designed to probe the problem–solving proficiency of the students.

4.3 Administration of the tests

Both tests which were printed into booklets were administered to the subjects in their respective schools by the authors. In the process, attempts were made to control some vital psychological factors concerned with test–taking. These incude motivation (Child, 1986) anxiety (Karmel, 1978), effort (Case, 1974). Test A was completed first, then test B.

4.4 Scoring the tests

In scoring test A, a grade of 1 (one) was awarded for selection of a correct option, while selection of a wrong option attracted a grade of zero thus giving a total grade of 40. In test B, a grade of 1 (one) was given for each operational step correctly executed and zero for each step wrongly carried out. Also, failure to attempt an execution of each and any of the steps attracted a grade of zero. Thus the total of maximum grade obtained in each problem depended on the number of operational steps required to successfully solve it. The maximum score for the test was 43. The steps involved in each of the free–response problems (i.e. skills being assessed) were operationally defined prior to scoring; i.e. at the time the test was being constructed. Also the knowledge-base required for executing such steps were conceptually defined during the formulation of test A.

As regards validity of the tests, the procedure suggested by Ebel and Frisbie (1986) for attaining intrinsic and rational content validity was adopted during the construction of the tests. In addition, the tests were submitted to a panel of 5 experts who were science educators for scrutiny and evaluation of their validity. Their comments, where necessary, were reflected in the final version of the tests.

ERIC

5. Results

The data were analysed using the "MINITAB" statistical package. The results are shown in Tables 1 – 5. Table 1 shows some of the statistical parameters of the scores.

Table 1: Some statistical properties of tests A and B

Statistical properties	Tests	
	A	B
Sample Size (N)	190	190
Sum of Scores (X)	3695	1643
Mean Score (X)	19.45	8.65
Std. Dev.	5.05	3.95

Tables 2a and 2b show the facility indices of tests A and B respectively, while Table 3 shows the results of a one–tailed t–test on the mean scores.

Table 2a: Facility indices of test A

F.I	No of items
> 0.50	20
> 0.30 ≤ 0.49	15
< 0.30	5
Total	40

Table 2b: Facility indices of test B

Problem No	F.I
1	0.44
2	0.21
3	0.28
4	0.16
5	0.18

Table 3: Comparison between students performances in tests A and B

Tests	Statistical Properties							
	N	$\bar{X}$	S	Std. Error of diff.	t–ratio	p	df	remark
A	190	19.45	5.09	0.47	22.98	0.05	378	significant
B	190	8.65	3.95					

Table 4 shows the Pearson's product–moment correlation coefficient (r) and the coefficient of determination (r^2) between the two scores.

Table 4: Association between tests A and B

Tests	Statistical properties				
	N	$\bar{X}$	S	r	r^2
A	190	19.45	5.05	0.45	0.20
B	190	8.65	3.95		

Table 5 shows the results of simple regression analysis of the two tests.

Table 5: Regression analysis of tests A and B

Tests	Statistical Properties					
	N	$\bar{X}$	S	No reaching predicted scores	% reaching predicted scores	coefficient of regression
A	190	19.45	5.05	80.00	42.00	0.35
B	190	8.65	3.95			

6. Discussion of the results and conclusions

In this study, the problem–solving capability of a sample of Nigerian pre–degree science students was investigated. The study sought to determine the strength of relatedness between the students' conceptual knowledge and their ability to extend such knowledge for use in solving contextual problems.

The conceptual knowledge measured by test A constituted the necessary prerequisite for solving the problems in test B. Ideally, a large percentage of the variance in test B should be accounted for by test A. Also, scores on test A should be good predictors of test B scores.

The relevant results are shown in Tables 1 – 5. Table 1 which shows the group mean scores and the respective standard deviations which indicate an average performance of 50 % for test A and only 20 % for test B. Table 2a shows that only 5 items out of the total of 40 have f.I lower than 0.30 whereas in test B, as shown in Table 2b all the questions except No.1, have f.I values less than 0.30. Table 3 shows that the students performed significantly better in test A than in test B. Table 4 shows that there is a positive, but low correlation ($r = 0.45$) between tests A and B. Also the coefficient of determination (r^2) between the two tests is 0.20. This shows that the variance in test B accounted for by test A is only 20 %. The result of simple regression analysis of the scores is shown in Table 5. The result obtained after entering the scores of each student in test A into fitted regression equation showed that only 80 out of 190, representing 42 % of the students obtained scores that were ≥ those predicted by their scores in test A.

All these results show that less than half of the students were able to apply the prerequisite conceptual knowledge that they possessed as shown by results of test A to solve the given set of problems in test B. These findings are in agreement with those of Perez and Terregrosa (1989) who obtained similar results in respect of Italian Secondary School students. It is also in agreement with the observation of Bunce et al (1991) that many of introductory chemistry students have difficulties solving chemistry problems.

The tentative conclusions from these results are as follows:

(a) The students possessed most of the pre–requisite conceptual knowledge needed for solving the problems.

(b) The students were not successful in solving the problems, not withstanding that they showed evidence of possession of the prerequisite conceptual knowledge.

ERIC Full Text Provided by ERIC

(c) There is a weak link between the students' possession of requisite conceptual knowledge and their problem–solving proficiency. The variance in the latter accounted for by the former was only 20 %.

(d) The view, often held by science teachers that students would develop desired problem–solving capabilities if only exposed to relevant conceptual knowledge should be taken cautiously especially where it forms a theoretical basis for instruction in science education.

7. Recommendations from the findings

The findings from this study which show that the students failed to solve the problems, even though they possessed most of the requisite conceptual knowledge can be interpreted in many ways. One of such interpretations, and which raises cause for concern is that learning has not taken place, even though the students may be scoring highly in achievement tests! According to Gagné (1985), the ability to solve problems is the hallmark of learning.

These results have a number of implications for science instruction generally and chemical education in particular. Some of these are outlined as follows:

(a) There is need for greater alertness and sensitivity, on the part of teachers (and science curriculum experts) to the imperatives of "teaching for understanding". For example, the structural organisation of curriculum materials and instruction should make it possible for a student who learns about "atomic structure" at some point in time to be able to use that knowledge in the understanding of "chemical reactions" of organic compounds in another segment of the programme. The excerpts of students' responses in test B, (appendix), indicate that such coherent conceptual pattern is lacking in their cognitive strategy.

(b) Teaching for understanding requires that the teacher, amongst other things, takes step from time to time, to locate students' conceptual difficulties which constitute stumbling blocks for problem–solving. In this way appropriate remediation can be administered. However, the over–crowded nature of the school syllabus stands in the way. This situation needs a review.

(c) Problem–solving heuristics should be specifically taught to students, not necessarily as additional layer of information, but rather as an element of instructional methodology. Studies have shown that explicit instruction in problem–solving strategies help to improve students' ability in general problem–solving. For example, Reif (1981), Chi et al (1981) Bunce et al (1991).

(d) The generalisation, which is often made by teachers and educators that students who "fail" tests or examinations lack the basic knowledge demands of such examinations seems to require a more careful analysis and clarification than hither to given. This study shows that such generalisations, and the assumptions which they engender should be approached with caution.

(e) The results highlight the need to appraise the scope and relevance of the aims of science teaching/learning in the school curriculum, to see if such aims are adequate and are being achieved. One way of doing this is through feedbacks from evaluation results. For example, students' inability to solve problems, as shown in this study can only be viewed as a pedagogic imperative if "problem-solving" is comprehensively defined as an aim of science education in the first instance. Hence a clear definition of curriculum aims vis-a-vis problem-solving skills is called for.

8. Suggestions for further studies

(a) The scope of generalisation of these results is somewhat, limited by the relatively few samples of school taken across the country. There is therefore, need to replicate the study, taking more schools as samples in order to justify the scope of confidence reposed on the generalisability of the results.

(b) It seems desirable to carry out similar investigations at the lower levels of the educational structure to determine whether students at these levels exhibit similar problem-solving behaviour as revealed by this study. Such results provide a rational basis for deciding the point at which the teaching of problem-solving skills can be introduced.

(c) The results of this study call for further investigations probably using interviews and analysis of resulting protocols to reveal the nature of conceptual bottle-necks that contributed to the students' inability to solve the problems.

References

Ajeyalemi, D. (1983). *The teaching of chemistry as experimental experience in Nigerian Secondary Schools: problems and prospects*. Journal of Science Teachers' Association of Nigeria, vol. 21 (2), 79 - 85.

Ashmore, A. D., Frazer, M. J. & Cassey, R. J. (1979). *Problem-solving and problem-solving networks in chemistry*. Journal of Chemical Education, vol. 56, 377 - 379.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York: Holt, Rinehart and Winston.

Butts, B. & Smith, R. (1987). *What do students perceive as difficult in H.S.C. chemistry?* Australian Science Teachers' Journal, vol. 32 (4), 45 - 51.

Bunce, D. M., Gabel, D. L. & Samuel, J. V. (1991). *Enhancing chemistry problem-solving achievement using problem categorization*. Journal of Research in Science Teaching, vol. 28 (6), 505 - 521.

Case, R. (1974). *Structures and strictures: Some functional limitations on the course of cognitive growth*. Cognitive psychology, vol. 6, 544 - 573.

Chi, M. H. H., Feltovich, P. J. & Glasser, R. (1981). *Categorization and representation of physics problems by experts and novices*. Cognitive science, vol. 5, 121- 152.

Child, D. (1986). *Psychology and the teacher*, 4th Ed. London: Holt, Rinehart and Winston.

Ebel. R. L. & Frisbie, D. A. (1986). *Essentials of Educational Measurement*, New Jersey: Englewood Cliffs, Prentice-Hall Inc.

Finley, F. N., Stewart, J. & Yarroch, W. L. (1982). *Teachers' perception of important and difficult* *science content*. Science Education, vol. 66 (4), 531 – 538.

Gabel, D. L., Sherwood, R. D. & Enochs, L. (1984) *Problem–solving skills of high school chemistry students*. Journal of Research in Science Teaching, vol. 21, 221 – 233.

Gagné, R. M. (1978). *Memory structure and learning outcomes*. Review of Educational Research, vol. 48, 187 – 222.

Gagné, R. M. (1985). *The conditions of learning*, 4th Ed. New York: Holt, Rinehart and Winston.

Garnett, P. J. & Treagust, D. F. (1992). *Conceptual difficulties experienced by senior high school students of electrochemistry, electric circuits and oxidation–reduction equations*. Journal of Research in Science Teaching, vol. 29 (2), 121 – 142.

Greeno, J. G. (1978). *Understanding and procedural knowledges in mathematics instruction*. Educational psychology, vol. 12 (3), 262 – 283.

Hackling, M. W. & Garnett, P. J. (1985). *Misconceptions of chemical equilibrium*. European Journal of Science Education, vol. 8 (4), 427 – 441.

Hayes, J. R. (1981). *The complete problem solver*. Philadelphia: The Franclin Institute Press.

Karmel, L. J. (1978). *Measurement and Evaluation in the schools*. 2nd Ed. New York: Macmillan.

Mitchell, I. J. & Gunstone, G. F. (1984). *Some student conceptions brought to the study of stoichiometry*. Research in Science Education, vol. 14, 78 – 88.

Perez, D. G. & Terregrosa, J. M. (1983). *A model for problem–solving in accordance with scientific methodology*. European Journal of Science Education, vol. 5, 447 – 455.

Reif, F. (1981). *Teaching problem–solving — a scientific approach*. The physics Teacher, vol. 19 (5), 310 – 316.

Robson, C. (1983). *Experimental design and statistics in psychology*. 2nd Ed. New York: Penguin Books Ltd.

Shaibu, A. A. M. (1988). *Topic difficulties in A–level chemistry: students' and teachers' perceptions*. School Science Review, vol. 70 (251), 118 – 120.

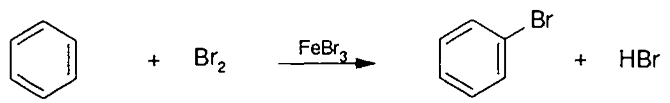
Stewart, J. (1982). *Two aspects of meaningful problem–solving in science*. Science Education, vol. 66. 731 – 749.

Appendix

1. Considering structural and solvent factors, outline the most probable reaction mechanism for the hydrolysis of bromopropane by aqueous potassium hydroxide. The equation of the reaction is:

$$CH_3CH_2CH_2Br + KOH\,(aq) \rightarrow CH_3CH_2CH_2OH + KBr$$

2. Outline the mechanisms of the reactions of bromine with benzene and with ethene; and briefly explain why a catalyst is needed for the first reaction but not with the second. The reactions are:

(I) 

(II) $CH_2{=}CH_2 + Br_2 \longrightarrow CH_2BrCH_2Br$

4. Using the relevant <u>mechanisms</u> of the respective reactions

(a) Explain why 1–bromobutane forms only <u>one</u> isomer, while 2–bromobutane forms <u>three</u> isomers respectively, when treated with an alcoholic solution of potassium hydroxide. The reactions are as follows:

$$CH_3CH_2CH_2CH_2Br \xrightarrow{\text{alc. KOH}} CH_3CH_2CH{=}CH_2 + H_2O + Br$$

$$CH_3CH_2CH(Br)CH_3 \xrightarrow{\text{alc. KOH}} \underset{\text{but-1-tene}}{CH_3CH_2CH{=}CH_2} + H_2O + KBr$$

$$\searrow_{\text{alc. KOH}} \underset{\text{but-2-tene}}{CH_3CH{=}CHCH_3} + H_2O + KBr$$

<u>Note</u>: Work out the third isomeric butene from reaction (II).

b) What name is usually given to the mechanism that you have described?

5. Consider reactions I and II below in which HBr adds across the carbon–carbon olefinic bonds:

(I) $\overset{1}{C}H_2{=}\overset{2}{C}H-\overset{3}{C}HO + HBr \longrightarrow \overset{1}{C}H_2Br-\overset{2}{C}H_2-\overset{3}{C}H{=}O$
nucleophilic addition

(II) $\overset{1}{C}H_2{=}\overset{2}{C}H-\overset{3}{C}H_3 + HBr \longrightarrow \overset{1}{C}H_3-\overset{2}{C}HBr-\overset{3}{C}H_3$
electrophilic addition

Outline the mechanisms of the reactions; and use it to explain why reaction II behaves differently from reaction I.

<u>Note</u>: –CHO is an electron withdrawing group.

COMMENTARY OF THE DISCUSSANT

Norbert Just, Universität Münster, Germany

Mr. Shaibu, recently you presented to us your study about the relationship between conceptual knowledge and problem–solving proficiency of Students in Nigeria. From this study you draw some conclusions which I will discuss in my statement.

At first I'm going to concentrate on the results of your study. You showed us, that students passes conceptual knowledge; but they are not successful in solving problems. You've found, that there is a weak relationship between the pre–requisite conceptual knowledge and the problem–solving proficiency. You've taken into consideration that these results are in agreement with different studies. In Germany the relationship between conceptual knowledge and problem–solving was investigated by Elke Sumfleth. The results of her study are similar to the facts you presented to us. Her conclusion is: Students know certain concepts but this does'nt say anything about their problem–solving capacity. They don't realise the context between their knowledge an the problem to solve. In this case the concept knowledge is isolated. You suggest for further studies to repeat your investigation and to generalise the results.

Therefore I would like to ask you: What do you expect of this replication and generalisation?

Another aspect in my opinion is your result that: for teaching understanding the overcrowded nature of school syllabus is a problem. Even in the last century a German chemistry expert, Rudolf Arendt, complained that the school–teaching of chemistry is often similar to the method of teaching at universities. His opinion is that students learn to many facts at school but the first aim of chemical education is to improve the students ability in logical thinking an problem solving.

Another chemistry teacher, Ferdinand Wilbrand, gives further arguments. He said in 1881: *"Der Unterricht in der Chemie soll den Lernenden mit den Methoden, Regeln und Hülfsmitteln der Induktion bekannt machen"*. (1881, p. 6) That means: Chemical education should teach students methods, rules and auxiliaries of induction. Not only heuristics has to be trained in school, but chemistry is considered as a method of learning this. Helmut Lindemann and Heinz Schmidkunz define problem–solving and discovering as the main task of chemical education. The structure of teaching and learning has to improve the capability of Students in problem–solving.

The curriculum of chemistry education leads students to the structure and hierarchy of the problem–solving–process. This process is basis for chemical education in this conception in schoolform.

My second question is: Wouldn't it be even more useful for further studies to include different curricula for chemical education to the study, respectively to do comparative investigations.

Finally: Mr. Shaibu, you define problem–solving as the first aim of science education and you call this a *"Pedagogic Imperativ"*. Therefore your definition is opposite to chemistry education by learning facts and concepts. The pedagogue Johann Friedrich Herbart distinguished in 1806 in his book *Allgemeine Pädagogik* between pedagogic and instruction. So we can ask whether the term "concept learning" means instruction and "problem–solving" education, or: what is the general sense of these terms.

If we give students in school some instructions about, for example, the structure of some substances, students learn facts. This process isn't educational if students have no chance to include their considerations about the why, the what and the way how to learn. The instructions habe to gain importance for the students.

Even more education is senseless if it basis only on the arrangement of the action–ability. Students can only improve their action–ability in a concrete coherence.

I think that in school instruction and education are two elements of learning, they assign both, concept learning and problem–solving their place in process of learning.

References:

Arendt, R. (1881). *Technik der Experimentalchemie*. Hamburg.

Herbart, J. F. (1986). *Systematische Pädagogik*. Hrsg. von Benner, D. Stuttgart.

Just, N. (1989). *Geschichte und Wissenschaftsstruktur der Chemiedidaktik*. Mülheim.

Schmidkunz, H. & Lindemann, H. (1992). *Das Forschend–entwickelnde Unterrichtsverfahren*. Essen: Westarp.

Sumfleth, E. (1988). *Lehr– und Lernprozesse im Chemieunterricht*. Frankfurt.

Wilbrand, F. (1981). *Über Ziel und Methode des chemischen Unterrichts*. Hildesheim.

SUMMARY OF THE PLENARY DISCUSSION

Holger Eybe, Universität Dortmund, Germany

The question of open and closed problems was raised in this discussion. Amos Shaibu used only closed problems, that is exercises, for the investigation. It was remarked that if a student is to solve a problem he/she has not met before then this could be regarded as an open problem. Thus, it may depend on the individual student whether a problem is open or closed. Also, a problem that appears to be closed to researchers and teachers may be an open problem for the students. Therefore, the mark scheme might be not flexible enough if students unexpectedly use strategies that deviate from the scheme. Amos Shaibu used a definition given in literature to categorise closed problems. He tried to make sure that the problems permitted only one solving strategy leading to the correct result.

As for error analysis and patterns in students' problem solving, it was pointed out that there was a second part of the investigation that involved think–aloud technique, taping, transcription and coding. However, as this reflects only a subjective point of view, patterns in problem solving cannot yet be derived.

It was pointed out that the exact correlation between test A and test B was not analysed. As the correlation coefficient is 0.45 the regression analysis may not be reliable.

ERIC

Analogies in Senior High School Chemistry Textbooks: A Critical Analysis

Rodney B. Thiele and David F. Treagust, Curtin University of Technology, Australia

1. Introduction

A currently supported view of learning is that when learners construct their own knowledge, it is both transferable to and usable in later learning situations. Recent research has shown that a significant factor enabling the creation of conditions where this type of learning occurs is related to teachers' subject matter understanding (Shulman, 1986; Kennedy, 1990). Of special importance is the teachers' content–specific pedagogical knowledge (Shulman, 1986). One aspect of this knowledge is the use of analogies which can effectively communicate concepts to students of particular backgrounds and prerequisite knowledge. Since students often lack the background to learn difficult and unfamiliar topics encountered in chemistry, one effective way to deal with this problem is for the teacher to provide a bridge between the unfamiliar concept and the knowledge which students have; this bridge can be provided by analogies. However, although analogies have been used in chemistry teaching in a variety of contexts (see for example, the work by Gabel and Sherwood, 1980), little research has been conducted in regular classroom settings about how chemistry teachers use analogies or how written materials which involve analogies are used by teachers and students. Consequently, there is need for research to investigate for whom and under what conditions analogies are most beneficial in chemical education.

In addressing this problem, a comprehensive research investigation is underway in chemistry education at the Science and Mathematics Education Centre at Curtin University of Technology to identify, examine and interpret those analogies used by authors of textbooks and by classroom teachers and their students at the secondary level. This research involves an examination of chemistry textbooks and interviews with the authors as well as observing and interviewing teachers and their students who are engaged in learning through the use of analogies in regular classroom settings. Similar studies are planned and also underway in physics and biology education.

To assist in the explaining of abstract chemical concepts, teachers and textbook authors may help students achieve conceptual understanding by employing teaching tools such as analogies. An analogy can allow new material to be more easily assimilated with the students' prior knowledge enabling those who do not readily think in abstract terms to develop a better understanding of the concept. Over the last ten years, heightened interest concerning the use of analogies in science education has resulted in the clarification of the picture of the types of analogies that are available and their range of presentation styles.

However, the use of analogies does not always produce the intended effects since some

ERIC

students take the analogy too far and are unable to separate it from the content being learned. Other students only remember the analogy and not the content under study whilst yet others focus upon extraneous aspects of the analogy to form spurious conclusions relating to the target content.

Recently, models have been developed which can be used as guide-lines for teachers and textbook authors concerning more effective analogy use and a classification system has been developed that enables researchers to systematically characterize textbook analogies. For example, both Glynn et al. (1989) and Zeitoun (1984) have presented model approaches to analogy teaching and the inclusion of analogies in textual materials. Whilst these reports are more recent than some of the textbooks under study, it is important to consider the possibility that the implications of these studies are not reaching textbook authors and curriculum designers. If this is the case, then we can expect that future editions of textbooks by these and other, authors will continue to present analogies in what some researchers (e. g. Curtis & Reigeluth, 1984; Duit, 1991; Glynn et al., 1989; Thiele & Treagust, 1991; Webb, 1985) consider to be a less than efficient manner.

This study involves a critical analysis of analogies found in chemistry textbooks used by Australian senior high school students, followed by an interpretative analysis of the views expressed by most of the authors of those books about the use of analogies in chemistry textbooks and chemistry teaching.

1.1 Defining an analogy

There is a need to clearly define what an analogy is so that it is not confused with illustrations, models and examples. For the purpose of this study, the researchers have adopted Glynn et al's. (1989) working definition:

"An analogy is a correspondence in some respects between concepts, principles, or formulas otherwise dissimilar. More precisely, it is a mapping between similar features of those concepts, principles and formulas." (p. 383)

The analogy requires the selection of a student world **analog** to assist in the explanation of the content specific **target** (or topic). The analog and target share **attributes** that allow for a relationship to be identified. Important in the presentation of a good analogy is some evidence of **mapping**. This process involves a systematic comparison of the corresponding analog and target attributes so that students are fully aware of which conclusions to draw concerning the target concept being addressed. It must be considered that both the analog and the target have many attributes that are not shared. Good mapping also gives some indication as to where this occurs so that unshared attributes are not ascribed to the target domain.

1.2 Different types of analogies

Reviews of analogy related literature (e. g. Duit, 1991) highlight a range of types of analogies which include verbal, pictorial, personal, bridging and multiple analogies, some of which are discussed below. Further, Curtis and Reigeluth (1984), in an analysis of 52 analogies from four American chemistry textbooks, proposed several other criteria by which analogies may be further classified by their integral parts. These criteria include an analysis of the nature of the shared attributes (structural or functional), the degree of explanation concerning the analog, as well as the level of enrichment of the analogy (the extent to which the author mapped the shared attributes). It is also evident that the final presentation by the classroom teacher will have a considerable influence upon the mode of operation of the analogy.

1.2.1 Verbal and pictorial analogies

Those analogies which include only written text or oral presentation are called verbal analogies. As this type of analogy is often subtly embedded in the body of the text, the reader is usually left to draw the necessary comparisons and conclusions about the target from the description of the analog if one is provided. Alternatively, a pictorial analogy, which includes some pictorial representation of the analog domain, allows the textbook author or teacher to pictorially highlight the desired attributes of the analog. This method helps provide a greater degree of visualization which reduces the likelihood that the student is not sufficiently familiar with the analog. Most pictorial analogies are accompanied by some verbal explanation and hence technically should be referred to as pictorial–verbal analogies.

1.2.2 Personal analogies

This type of analogy is believed to assist students by relating abstract chemical concepts to student world phenomena such as people, money, food and relationships. For example, the text may encourage the students to imagine that they are packaging sausages and rissoles into barbecue packs with each pack containing exactly two sausages and one rissole. This may be shown to be analogous to a stoichiometrically reacting system and the effect of a limiting reagent on the amount of product and excess reagent remaining in the system. Marshall (1984) suggests that this type of analogy causes better learning of concepts and that the approach is more enjoyable although she warns that personal analogies can cause students to give intuitive feelings to inanimate objects and concepts.

1.3 The advantages of analogies in teaching

Analogies are believed to help in three major ways in that they: a) provide visualization of abstract concepts; b) help compare similarities of the students' real world with the new concepts; and c) have a motivational function.

1.3.1 Visualization process

Researchers (Glynn et al., 1989; Shapiro, 1985) agree that the visualization process is very important in the learning of concepts and that the analogies prompt a visualization process to aid understanding. In an analysis of 216 analogies found in science textbooks for secondary students, Curtis and Reigeluth (1984) found that chemistry textbooks contained the highest percentage of pictorial analogies (29%) compared to the total science average of only 16%.

1.3.2 Real world linkage

It has been proposed that analogies have been historically linked to both the explaining of science and to the processes of science. Well renowned theorists such as Maxwell, Rutherford and Einstein are reported to have used analogical reasoning as a tool to aid problem solving, to explain hypotheses and to communicate to audiences about early theories of atomic structure (Lewis & Slade, 1981; Shapiro, 1985). Needless to say, analogies are used more frequently when the target domain is difficult to understand or is foreign to the learner (Duit, 1991). The presentation of a concrete analog in this situation facilitates understanding of the abstract concept by pointing to the similarities between objects or events in the learners' world and the phenomenon under discussion.

1.3.3 Motivational function

The motivational sense of analogy is due to a number of factors. As the teacher or textbook author is drawing from the students' real world experience, a sense of intrinsic interest is generated. In addition to this interest, students who traditionally perform at lower academic levels may be more likely to achieve a greater level of conceptual understanding. This should result in a motivational gain for the students. However, it should be noted that little has been determined from empirical studies about the actual learning processes that are associated with analogy assisted instruction since most of the studies have measured only the students' recall of learned materials. It is also not well known if analogies really do assist students to attain a level of conceptual understanding or whether students only use the analogy as another algorithmic method to obtain the correct answer.

1.4 The constraints of analogies

Despite the positive outcomes of analogies stated above, the use of analogies as a teaching tool can cause incorrect or impaired learning due to several fundamental constraints related to the analog — target relationship. Three of these constraints are analog unfamiliarity, the student's cognitive development and the incorrect transfer of attributes.

1.4.1 Analog unfamiliarity

A significant constraint on the use of analogies in teaching is the learner's unfamiliarity with the analog selected by the textbook author or teacher. Several empirical studies on the use of analogical reasoning in chemistry instruction, for example studies by Gabel and Sherwood (1980, 1983, 1984), indicated that a significant proportion of students did not understand the analog sufficiently well. These results emphasise the need for caution in teaching with this method and in making instructional decisions based on an evaluation of those analogies that are presented to improve student understanding of chemistry concepts. A strategy that can be employed by textbook authors to reduce the problem of analog unfamiliarity is to provide additional 'analog explanation' concerning the analog and its relevant attributes. This will provide useful, additional information to the student. 'Strategy identification', where the author engages a term such as "analogy" or "analogous", may serve as a warning to students that careful thought is required to derive the full and correct meaning from the analogical statement.

1.4.2 Student's cognitive development

A second area of constraint with analogy use relates to the Piagetian stages of cognitive development. There is general agreement that analogies can assist students who primarily function at lower cognitive stages; however, if these students lack visual imagery, analogical reasoning, or correlational reasoning, the use of analogies is still believed to be limited (Gabel & Sherwood, 1980). In addition, those students already functioning at a formal operational level may have already attained an adequate understanding of the target and the inclusion of an analogy might add unnecessary information loads that could also result in new misconceptions being formed by the students.

1.4.3 Incorrect transfer of attributes

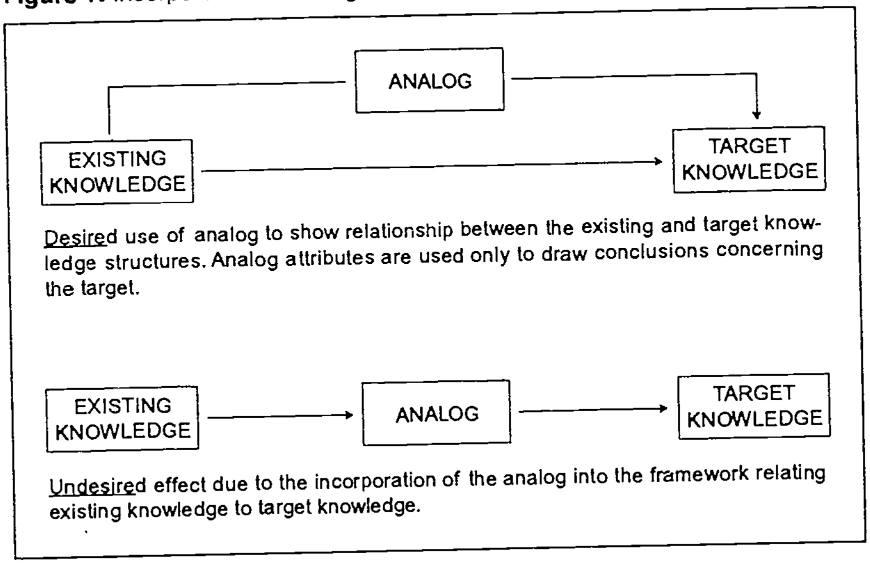
The nature of the analog is that it has some shared attribute(s) with the target. However, it may be considered that the unshared attributes are as instructive to the students as are the shared attributes (Licata, 1988). No analog shares all its attributes with the target as, if it did, the analogy would then become simply an 'example'. These attributes that are not shared are often a cause of misunderstanding for the learners if they attempt to transfer them from the analog to the target.

Textbook authors may provide further mapping in an attempt to reduce the likelihood of the student incorrectly transferring analog attributes. Curtis and Reigeluth (1984) have reported that all of the 52 analogies found in four popular American chemistry textbooks included some statement concerning the nature of the shared attributes although only 10 analogies had mapped more than one attribute to create an 'extended' analogy. This encouraging lack of 'simple' analogies was not, however, characteristic of science textbooks in general.

A related constraint occurs when the students attempt attribute transfer in an inappropriate manner. Rather than using the analog attributes as a guide for drawing conclusions concerning the target, the students occasionally incorporate parts, or all, of the analog structure into the target content. This is illustrated in Figure 1 on the next page. One of the results of this incorrect transfer is that when students are questioned concerning the nature of the target content, they will answer with direct reference to the analog features. For these reasons, some instructors choose not to use analogies at all and thereby avoid these problems whilst forsaking the advantages of analogy use.

When analogies are used during classroom instruction, discussion should take place to assist in the delineation of boundaries and to aid concept refinement (Licata, 1988; Webb, 1985). Indeed, Glynn et al. (1989) have produced a six step Teaching–With–Analogies model that is designed to assist teachers use analogies effectively. This model provides for a clear delineation of shared and unshared attributes by the teacher. Allowing for student involvement and discussion at the classroom level also will provide feedback to the instructor if incorrect attribute transfer has occurred. Teachers should not assume that the students are capable of effecting correct analogical transfer unassisted but, rather, they should provide explicit instruction on how to use analogies and provide opportunity for classroom discussion on the subject. The highlighting of unshared attributes may also be done by a textbook author although documented evidence of this occur-

Figure 1: Incorporation of analog in new knowledge.



rence is rare (Curtis & Reigeluth, 1984). It may be that textbook authors assume that this is not required, or that teachers will conduct this aspect with the students in class time. Recent research, however, indicates that teachers do not expand upon analogies contained in their students' textbooks when conducting their normal teaching routines (Treagust, Duit, Joslin & Lindauer, in press).

2. Research focus

This study was conducted in two parts. Firstly, an examination of ten chemistry textbooks used in Australian senior high school chemistry classrooms was carried out in order to determine the extent and nature of the analogies. Specifically, the study investigated the frequency of analogies found in these textbooks; compared the frequency of analogy use for particular sections of the subject matter or at different stages of the curriculum; identified textbook authors' incorporation of instructional strategies that aim to directly assist the student to use analogies to aid understanding; and examined the type of analogies used most frequently in the textbooks.

Secondly, interviews were conducted with the authors of seven of the ten textbooks. Specifically, this part of the study solicited the views currently held by the textbook authors concerning analogy use; examined authors' reasons for inclusion or exclusion of analogies in instructional materials; any personal appeal for a model approach to analogy teaching; and investigated the changes the author would make to a later edition of their own textbook if they were provided with a more thorough repertoire of trialled, familiar analogies.

The two parts of this study are described separately followed by conclusions drawn from the results of both parts and implications for teachers, textbook authors and educational researchers.

3. Analogies used in chemistry textbooks

3.1 Procedure

Ten chemistry textbooks (see Appendix) were examined and all analogies identified were photocopied and further analysed. The textbooks used in the analysis had been identified by state syllabus organisations as those current, generally used textbooks for Australian senior secondary chemistry education. Only one of the textbooks, a British publication, was not published in Australia.

A portion of text or a picture was considered to be analogical if it was aligned with the working definition stated above (Glynn et al., 1989, p. 383) and/or it was identified by the author as being analogical. Each analogy was scrutinized concerning the following features, three of which (c, d and e) were reported for American science textbooks by Curtis and Reigeluth (1984): (a) the content of the target concept; (b) the location of the analogy in the textbook; (c) whether it was verbal or pictorial; (d) evidence of further analog explanation or strategy identification; (e) the extent of the mapping done by the author; and (f) the presence of any stated limitation or warning.

3.2 Results

A total of 93 analogies were identified from the ten textbooks. This resulting mean of 9.3 analogies per textbook is less than the mean of 13 reported in the American study (Curtis & Reigeluth, 1984). The number of analogies found in each book varied considerably with five books having less than six analogies whilst the other five had between 12 and 18 analogies. Each analogy was further classified independently by the two researchers with an original agreement of 93 %. The remaining 7 % of the classifications were agreed upon following consensus discussions.

3.2.1 Content analysis

The content area of the target concepts were classified into 15 categories. Table 1 indicates that a considerable proportion of the analogies (21, 22.6 %) relate to "Atomic Structure" — including electronic arrangement. Other areas in which analogies were used more frequently were found to be "Energy" — including collision theory — (11, 11.8 %) and "Bonding" (12, 12.9 %). The submicroscopic nature of these target concepts emphasizes the visualization role of analogies.

Table 1: Analysis of the frequency of analogy use compared to target content area.

Content area	n	%	Content area	n	%
Acids & Bases	6	6.5	Industrial Processes	1	1.1
Analytical Methods	3	3.2	Nature of Matter	8	8.6
Atomic Structure	21	22.6	Organic Chemistry	5	5.4
Biochemistry	6	6.5	Periodic Table	2	2.2
Bonding	12	12.9	Reaction Rates	3	3.2
Chemical Equilibrium	5	5.4	Solutions	3	3.2
Chemical Processes	1	1.1	Stoichiometry	6	6,5
Energy	11	11.8			

4.3.2 Analogy location in textbook

The page number of each analogy was used to determine a decile measure of the analogy's location within the textbook as a whole. Table 2 provides data which suggest that the analogies tend to be located more frequently in the earlier stages of the textbook except for a number found in the 7th decile. This could indicate that conceptual targets are encountered in two phases — initially when the new work is being introduced and also, at a later phase, when more difficult concepts are being presented.

4.3.3 Verbal and pictorial analogies

Forty four (47.3 %) of the identified analogies had a pictorial component. These pictorial analogies included some diagrammatic representation of the analog. Further analysis revealed that pictorial analogies are frequently positioned in the margin, presumably as an anecdotal package of helpful information. However, as Table 3 illustrates, verbal analogies were rarely found in a marginalized position which indicates that authors may wish to use pictorial analogies more frequently but tend not to sacrifice the copy space. Those authors writing texts with marginalized comments tend to make use of the opportunity to use this space for pictorial analogies.

Table 2: Analysis of the decile position of the analogies in the textbooks as a whole

Location	n	%	Cum %
0	21	22.6	22.6
1	12	12.9	35.5
2	14	15.1	50.5
3	9	9.7	60.2
4	9	9.7	69.9
5	4	4.3	74.2
6	9	9.7	83.9
7	12	12.9	96.8
8	3	3.2	100.0
9	0	0.0	100.0

Table 3: The frequency of use of marginalized and pictorial analogies in the textbooks

	Marginalized	Text Body	Total
Verbal	5	44	49
Pictorial	25	19	44
Total	30	63	93

3.2.4 Further analog explanation or strategy identification

To avoid the problems of analog unfamiliarity and incorrect attribute transfer, some writers provided background information concerning the relevant attributes of the target domain. This analog explanation attempts to ensure that the student is focussing upon the appropriate attributes at the time of analogical transfer. The explanation may constitute a simple phrase of only a few words through to a paragraph thoroughly describing the analogy attributes: 56 (60.2 %) of the analogies had some analog explanation

ERIC

which is a little lower than other reported research (Curtis & Reigeluth, 1984).

Further, only 15 (16.1 %) of the analogies included any statement identifying the strategy such as "an analogy", "analog", or "analogous". It is likely that if 'strategy identification' was employed more frequently, then the effect would be similar to the addition of a warning in that it will direct students towards the correct cognitive procedure and the students may be less likely to transfer analog attributes incorrectly to the target (Glynn et al., 1989).

3.2.5 The extent of mapping

The extent of mapping used by the textbook authors was classified using Curtis and Reigeluth's criteria of 'Level of Enrichment': (a) simple — states only "target" is like "analog" with no further explanation; (b) enriched — indicates some statement of the shared attributes; and; (c) extended — involves several analogs or several attributes of one analog used to describe the target.

The textbook analysis found that the use of simple analogies was fairly common (42, 45.2 %) despite research suggesting that students require assistance when relating the correct analog attributes to the target. This figure is substantially greater than that reported by Curtis and Reigeluth for chemistry textbooks. Only 35 (37.6 %) of the analogies were enriched whilst the remainder (16, 17.2 %) were extended. Further, three of the six textbooks having 12 or more analogies contained considerably more simple than enriched analogies.

3.2.6 Limitations

Given that analogies can be used incorrectly by students and that research suggests that authors include some warning as to the limitations of the analogical process, each analogy was examined to see if it included either a general statement of the limitation of analogy use or a statement relating specifically to the unshared attributes in that analogy. No general statements concerning analogy use were made in any of the textbooks. and only eight specific warnings or limitations were expressed. These data add support to the suggestion that authors are either assuming students are capable of effecting the analogical transfer themselves or that the teacher, in the course of normal classroom teaching, assists in this regard.

3.3 Discussion

While analogies were used slightly less than those in the U.S. study, specific areas of chemistry subject matter, characterized by their microscopic processes and abstract nature, used analogies most. In addition, the analogies tended to be more common in the early sections of the textbooks where prerequisite concepts are being established and authors may be more likely to engage student–friendly strategies. The link between visualization processes and analogy is borne out by the significant use of pictorial analogies. These were often positioned in the page margin beside the text which may indicate that one of the reasons why authors resist using more pictorial analogies and verbal analogies also for that matter, is due to a lack of copy space.

Often, the authors employed further analog explanation and this must be encouraging to researchers although evidence of strategy identification was scarce. In addition, the number of simple (unmapped) analogies was found to be much higher than that reported in the U.S. study. This possibly indicates the authors' impression that the classroom teachers will provide further support to the analogy for the students. Similarly, little evidence was found of genuine attempts by authors to highlight the limitations of a stated analogy. This special form of mapping may also require the assistance of the classroom teacher if the analogy is to prove most effective.

4. Authors' perspectives on analogy use in chemistry textbooks

4.1 Procedure

Semi-structured interviews were conducted with seven authors referred to as A, B, C, D, E, F and G who represent eight of the ten textbooks analysed. Of the remaining two textbooks, one author was no longer in a position to be involved due to failing health and the other was overseas.

The interviews were conducted in a semi-structured manner as suggested by Yin (1989). Six of the seven interviews were conducted in person and lasted between 60 to 80 minutes. The other interview (with Author E) was conducted via a long distance telephone conversation which was tape recorded with that author's permission. During each interview, examples of analogies identified from the author's own textbook were used, wherever possible, to focus discussion and to assist in the definition of terms used by the interviewer and interviewees. All interviews were tape recorded and transcribed and subsequently the transcripts were analysed in an interpretative manner (Erickson, 1986) to address the research focus. At this stage, further reflective observations from the transcripts were provided by a colleague.

The results of the textbook analysis indicated that Authors A, E and G used analogies more frequently (12, 14 and 14 analogies respectively) whilst the textbooks represented by the other four authors contained between one and five analogies. However, the textbook by Author E contained almost four times as many words as the other textbooks and, on an analogy by word count analysis, can be considered to contain a similar number of analogies to the textbooks by Authors B, C, D and F.

4.2 Results

4.2.1 The characteristics of a good chemistry teacher

Each author was asked to briefly describe the characteristics of a good chemistry teacher to ascertain any general leaning towards pedagogical styles that would be particularly conducive or otherwise towards the use of analogies in textbooks.

Five of the authors (B, C, D, E and F) strongly emphasized that the need for a strong background in chemistry was by far the most important factor. Comments, such as that of Author E, about the requirement of a teacher to be "totally on top of the discipline

ERIC

aspect" and by Author D, of the need for a content knowledge "way in advance of the level you are teaching" as being "the only way that you can have comfortableness, enthusiasm, imaginative ideas, different ways of presenting things" suggests that these authors consider that good teaching needs a foundation of good content knowledge. Having established this foundation, the authors stated that being interested in students and being able to suitably organise and select the content material were other characteristics of good teaching. For example, having commented upon the need for chemistry teachers to know their subject matter, Author F indicated that "secondly, you have to teach where the students start. You have to know where their knowledge is and what their interests are and start from that basis so that you can build on something." However, two authors (A and G) considered student–teacher relationships as the most important characteristics of good chemistry teaching. Author A stressed the need for the teacher to be interested in the thought processes of students, and to be "clear and precise, be fair, and be prepared to admit that you are wrong" whilst Author G proposed that "it's really mostly important just to get the students interested".

4.2.2 The meaning of analogy

The authors' ideas of what did, and what did not, constitute an analogy, indicated general agreement with each other and with the research literature. Some variances in the discussion, however, followed the lines of "all science is analogy" due to the use of symbols in instruction and descriptions of invisible processes and entities (Author C). One of the authors (D) discussed and demonstrated what could be described as a rice analogy for particle theory relating to the states of matter (Knox & George, 1990). It was agreed that this demonstration, although being analogical in nature, could better be described as an analog model. The authors demonstrated that they did not confuse examples with analogies and they were able to clearly identify two discrete domains in the analogies that were discussed although, for several of the authors, there was evidence of a lack of delineation between "analogy" and "model".

4.2.3 Examples of analogies used in their own teaching

Each author was asked to recount several analogies used in their own teaching in order to provide insight into the authors' personal views of analogies in teaching so that a comparison could be made to the frequency of analogy inclusion in the textbook.

Four of the authors (A, B, C and D) conveyed some difficulty recalling analogies that they had used in their teaching. Author E was able to instantly respond with an analogy that he had recently used in a teaching situation; however, this author had some prior warning of the questions due to the nature of the telephone interview arrangements. Author B suggested that he did not use analogies as frequently now as what he had done at the time when he was teaching chemistry using the Chem Study curricular materials and, in explanation, he commented that he was well aware of the research indicating that some of these analogies may result in the students forming misconceptions.

One of the authors (A) indicated several analogies that were in his textbook that he used in his teaching also. He suggested that the analogies lost something when they went to print because, when using them in the classroom, they could be presented in such a way as to foster interest and motivation more readily. He, like the others, indicated that analogy was something of a spontaneous exercise and that they were more likely to use analogies when attempting to explain an abstract idea to students after the students had indicated that they did not clearly understand. For example, Author C commented that "analogy is a very personal thing, something you might deal with on a one to one basis." Despite freely acknowledging that he used analogies in his own chemistry teaching, Author F also commented upon the need to change or adapt models and analogies to suit the changing circumstances of the lesson and pupils.

In a similar manner, all of the authors seemed to be aware that there was a need for analogies to be discussed by both teachers and students when they are used in a classroom setting. Having described his analogy for the semi–conductor, Author E commented that the analogy had been created spontaneously by him as a result of being questioned concerning the target concept by some students after a lecture. In this situation, he could "push it [the analogy] to outrageous lengths, then, when they've seen the point, you can, sort of, throw the analogy away and come back to the point you are trying to make." Author C suggested that when using analogies on the whole class scale, that he would build the analogy and then destroy it to illustrate where the analogy broke down. He proposed that the instructional value was in the resulting discussion and evaluation of the unshared attributes rather than in the construction or presentation of the analogy. Author B, when questioned concerning a particularly problematic analogy for chemical equilibrium commented that "maybe the way to deal with it is to point out what's wrong with the model." Later, with respect to the same analogy, he suggested that "I wouldn't mind using it myself if I had control of the situation in a classroom situation" but he went on to indicate the lack of control available once the analogy is in textbook form by adding "but I wouldn't like to stick it in a text where everybody is going to use it."

This theme of the inability to negotiate analogy once it is in text was taken up by several authors. For example, Author A commented that "many analogies work much better in the teaching situation than they do as presented in books." Author B felt likewise and suggested that he preferred to use them himself "rather than put them in written form .... I don't really want to lock an analogy into concrete." Author E, commenting from a similar viewpoint, proposed that "there is a bit of a danger that students can get the analogy confused with the reality". Similarly, Author C responded that "I would be reluctant to use an analogy in a textbook because I don't know that you could ever, in words, provide an adequate representation for all students." Later, he emphasized that when you are teaching you can respond to those students who do not understand the concept by an analogy or further explanation but that "you don't see the blank faces of the students you are writing the text for."

One important feature that arose was the degree to which the authors anticipated that the classroom teachers would engage themselves in explaining the content of the textbooks to their students. For example, when asked about the need to identify analogies found in the textbook with some strategy identification, Author F remarked that he didn't know if they'd described it as an analogy in the textbook but that "certainly I would teach it that way... I don't think we put teaching techniques in the textbook." This author went on to describe an example of how he taught using analogies directly from the textbook, highlighting things that the two domains had in common as well as any limitations. Author G also was able to describe how she taught directly from the textbook and discussed the analogies with the students in the classroom. However, that author remarked that, whilst "it was the teachers' role to explain [what was in the text-

ERIC

book], ... if the student was away, the textbook should be sufficient instruction". Further, Author B indicated that the book was clear enough that students should be able to work through it themselves without undue difficulty. This could indicate that textbook authors do not presuppose that teachers, in normal circumstances, will explain analogies and analogy limitations to the students in many cases.

4.2.4 Students' understanding of analogies

Often, comments were made by the authors concerning the interactions that the students might have with the analogy and the process of analogy. When asked why he decided to include a particular analogy in his text, Author D responded in terms of traditionally accepted advantages for analogy inclusion by stating that "we use an analogy that is in our experience, that is ... something we can relate to and can visualize." Author C remarked that we use analogy "to help put something into a language that the students can understand ... to make the complex commonplace" whilst Authors F and G referred to analogies "relating to real life activities" and being "something familiar" respectively.

These authors have identified the 'visualization' advantage and the 'student world' advantage, yet not the 'motivational' advantage of analogy. It is also interesting that Author D considered that the analogy should be in "our" experience and as being something that "we" can relate to. This is important when compared to the comments of Author C who, reflecting back on some research he conducted in classrooms, recalled a student making a statement to the effect that:

"Analogies are very personal things. What's a good analogy for Mr X is not a good analogy for me because I don't think the same way that Mr X does. It's alright for Mr X — he knows the whole story. I don't! You're trying to present an analogy to me when you know what all of it means. You know what is coming up and you are aware of the history behind that. Here I am, as a student in the first couple of weeks of my senior chemistry course, being thrown into this same thing. I don't know what the end of the story is like. It's like trying to use information that I am going to get in Chapters 7, 8 and 9 of my novel to answer a dilemma that I have in Chapter 1."

Author C also acknowledged that, whilst the students' ability to deal with the analogy should not be overlooked, the need for analogies varied markedly from student to student. He believed that the need is related more to academic ability and suggested that "there have got to be some students who will need a little bit more information and there will be other students who will say 'this is pretty tedious stuff'." In a related discussion, Author D argued that

"You put it off to the side because it's something you could elude to and is maybe useful but it doesn't interfere with the flow and it's not necessary for the flow and not everyone would need it but it's there if you want to go that way."

4.2.5 Awareness of, and appeal for, a model approach to analogy teaching

With this question, the researchers wished to ascertain which, if any, of the authors recognized Glynn et al's Teaching-With-Analogies model or were aware of this or any other models relating to analogy presentation. In addition, the authors were asked to comment upon their perception of the usefulness of such a model for textbook presentation of analogies after having studied it for several minutes.

None of the authors recalled ever having seen or used a model for analogy presentation. One of the authors (C) was aware of the work done by Clement on bridging analogies in physics (Clement, 1987) and Author B commented upon the similarity of the model to an established approach for the teaching of concepts (with attribute analysis, examples and non–examples). Author E was content with the model as it was presented and indicated that it was common sense and that "most experienced teachers would do that without even being conscious that they were doing these things." The other authors did appear to accept the six step model as useful although some suggested variations and alterations, while others (A, D, F and G) acknowledged that they considered that the best approach varied depending upon the analogy and the setting. The problem of extra length to a textbook was raised by Author F who proposed that analogies in this model approach could be better placed in a teachers' guide to the textbook. Author G cautioned that, whilst there was nothing wrong with the model, he would "never like to see a recipe for how you teach."

Other useful comments came from Author B who suggested that the limitations of an analogy could be presented at the same time as the similarities — that is, before the conclusions are drawn. Author D attempted to clarify the conflict that arises over the use of an analogy to draw conclusions rather than to confirm a previously arrived at conclusion by remarking that:

"There certainly must be cases where the conclusion has already been made and, in an attempt to make sense of that conclusion, you might use the analogy. Though, I suppose what you are at least doing is drawing out the conclusions of what you have previously done before."

The authors' comments indicate a general acceptance of the model approach for analogy presentation provided that due regard be given for flexibility in various settings.

4.2.6 Proposed future use of a bank of trialled analogies

When asked if they would be interested in incorporating some of a bank of trialled analogies in a fictitious new edition of their texts, several of the authors indicated some reluctance. Author C clearly stated that he "wouldn't use them in the textbook" whilst Author B only agreed to the inclusion on the proviso that the "person writing the book felt comfortable." Similarly, Author E indicated that he would consider them but he would "need a fair bit of convincing that an analogy was a valid one." Authors of the more recent texts (Authors F and G) indicated little willingness to deviate from their current texts. On further questioning, however, four authors showed particular interest in having these trialled analogies available in the form of a teachers' guide. Author E showed enthusiasm for the idea whilst Author D proposed that each analogy "could have a blurb on each of the six steps". In addition, Author C suggested that he "might be interested in putting them in a teachers' guide." He qualified his concession, however, by proposing that, even in a teachers' guide, there would be the need for some form of instruction to teachers that "clearly you don't just take these into the classroom cold and assume that all your students, at the end, will be enlightened." Alternatively, Author F suggested that although you could embed these analogies into another chemistry book, there would be demand for a "book of analogies.... that's related to a number of chemical concepts that are pretty universal".

ERIC

4.3 Discussion

This part of the study has highlighted differences in the authors' approach to analogy that may explain the variations in the frequency of analogy inclusion in the textbooks examined in the first part of this study. Generally the authors felt uncomfortable about setting analogies into print as the sense of control and flexibility by the teacher is lost. Most consider that analogies should be used by teachers as a negotiated response strategy to be used when the teacher considers that an explanation to the student/s about particular concepts have been unsuccessful.

The author's reasons for including analogies in their textbooks coincided with two of the main advantages of analogies as reported in the literature — namely, the provision of a 'student world' view and the improvement of the visualization process; however, they did not directly recognise the reported advantage of improving student motivation. Those authors who have used analogies infrequently in their textbooks tended to place more importance upon the teachers' subject matter knowledge than upon their maintaining students' interest and developing relationships. A further factor that may have limited the use of analogies was the space that they require in the textbook. Most of the authors made incidental remarks, during the interviews, of the pressure that they were under by publishers to keep the textbook to the smallest possible size and price.

The authors were not familiar with any models that guided analogy presentation in textbooks or in teaching. However, upon perusal of such a model, most of the authors suggested that it could be useful although they expressed a desire to have some flexibility in the order of the stages of the model depending upon the teaching situation. The interviews with the authors showed that the use of analogy in the page margins of some textbooks reflects the perception of the author that the analogy is something that many students can do without and should, therefore, simply be made available to those students requiring a further explanation of the concept under study.

Those authors who had used analogies sparingly in their textbooks showed little or no inclination to include them in a later, fictitious textbook. One author, who used analogies frequently in both the textbook and in teaching, suggested that there would be no change in the future edition. The authors expressed interest in the inclusion of a bank of trialled analogies into a teachers' guides where an approach such as the Teaching–with–Analogies model could be used to assist teachers in their teaching with analogies. Support for a resource book of chemistry analogies was also voiced.

5. Conclusions

Research into how students and teachers use analogies in the learning/teaching enterprise indicates that enriched analogies, rather than simple analogies for all but the most elementary relationships, increase the effectiveness of analogical transfer and hence, the understanding of the target domain. Similarly, research suggests that analogies used in textbooks where there is a lack of instruction or assistance in using the analogical processes and a scarcity of stated limitations, are less useful than when these features are included. In analysis of the ten textbooks reported in this study and the interviews with the authors, both these issues would appear to require greater attention in order to optimise learning of concepts by analogy use. The authors of the textbooks have as-

ERIC

sumed that the classroom teacher will use the analogies in such a manner to enhance their pedagogical use. However, there is no research evidence to support this suggestion since teachers' pedagogical content knowledge has been known not to be generally extensive in this regard (Shulman, 1986; Treagust, et al., in press) and the interviews with the authors provided no recommendations of what teachers were expected to do with the analogies and, with the exception of two textbooks, teachers' guides were not written for the textbooks.

Glynn et al. (1989) described analogies as double–edged swords and the chemistry textbook authors would in many instances identify with this perception. However, the inclusion of analogies in the textbooks written by these authors reflects variations in the authors' perceptions of the advantages and disadvantages of analogy use as well as reflecting variations in their respective teaching backgrounds. The frequency of analogy inclusion in the textbooks does not seem to be indicative of the willingness of the authors to use analogy in their own teaching; rather they are unwilling to set the analogy to print because of their belief that teaching with analogies should involve discussion or negotiation with the students. This is not possible in a textbook situation.

The unfamiliarity of the authors with research guides regarding analogy presentation highlights the problems of the efficient dissemination of research findings in science education to practitioners. However, the willingness of the authors to accept a model approach to analogy, albeit a more flexible one, indicates the usefulness of such an approach. Also, it should be considered that there is still a lack of empirical research findings suggesting that analogy presented in the model format aids student understanding more than other analogies.

As we observe chemistry teachers and their students in regular classroom settings using analogies to better understand complex concepts, we anticipate that we will be able to determine for whom and under what conditions analogies are most beneficial in secondary school chemistry. Subsequently, we plan to provide materials, in the form of a teachers' guide, which is consistent with the recommendations of the textbook authors and which will enable analogies to be used by chemistry teachers and their students in an exemplary fashion.

References

Clement, J. (1987). *Overcoming misconceptions in physics: The role of anchoring intuition and analogical validity*. In: J. Novak (Ed.), Proceedings of the 2nd International Seminar on Misconceptions and Educational Strategies in Science and Mathematics, vol. III, pp. 84 – 97. Ithaca, NY: Cornell University.

Curtis, R. V. & Reigeluth, C. M. (1984). *The use of analogies in written text*. Instructional Science, vol. 13, 99 – 117.

Duit, R. (1991). *On the role of analogies and metaphors in learning science*. Science Education, vol. 75, 649 – 672

Erickson, F. (1986). *Qualitative methods in research on teaching*. In: M. C. Wittrock (Ed.), Handbook of Research on Teaching (3rd ed.), pp. 119 – 161. New York: Macmillan.

Gabel, D. L. & Sherwood, R. D. (1980). *Effect of using analogies on chemistry achievement according to Piagetian level*. Science Education, vol. 64, 709 – 716.

ERIC

Gabel, D. L. & Sherwood, R. D. (1983). *Facilitating problem solving in high school chemistry.* Journal of Research in Science Teaching, vol. 20, 163 – 177.

Gabel, D. L. & Sherwood, R. D. (1984). *Analyzing difficulties with mole-concept tasks by using familiar analog tasks.* Journal of Research in Science Teaching, vol. 21, 843 – 851.

Glynn, S. M., Britton, B. K., Semrud-Clikeman, M. & Muth, K. D. (1989). *Analogical reasoning and problem solving in textbooks.* In: J. A. Glover, R. R. Ronning & C. R. Reynolds (Eds.), Handbook of creativity: Assessment, theory and research (pp. 383 – 398). New York: Plenum.

Kennedy, M. M. (1990). *A survey of recent literature on teachers' subject matter knowledge.* Issue Paper 90 – 3, The National Centre for Research on Teacher Education, Michigan State University, East Lansing, Michigan.

Knox, K. & George, A. (1990). *Close packing — A dynamic demonstration.* Journal of Chemical Education, vol. 67, 700 – 701.

Lewis, P. & Slade, R. (1981). *A guide to H.S.C. chemistry.* Melbourne: Longman Cheshire.

Licata, K. P. (1988). *Chemistry is like a ...* The Science Teacher, vol. 55 (8), 41 – 43.

Marshall, J. K. (1984). *Classroom potpourri.* Journal of Chemical Education, vol. 61, 425 – 27.

Shapiro, M. A. (1985). *Analogies, visualization and mental processing of science stories.* Paper presented to the Information Systems Division of the International Communication Association, Honolulu, HI.

Shulman, L. S. (1986). *Those who understand; knowledge growth in teaching.* Educational Researcher, vol. 15 (2), 4 – 14.

Thiele, R. B. & Treagust, D. F. (1991). *Using analogies in secondary chemistry teaching.* The Australian Science Teachers Journal, vol. 37 (2), 10 – 14.

Treagust, D. F., Duit, R., Joslin, P. & Lindauer, I. (in press). *Science teachers' use of analogies: Observations from classroom practice.* International Journal of Science Education.

Treagust, D. F., Duit, R., Lindauer, I. & Joslin, P. (1989). *Teachers' use of analogies in their regular teaching routines.* Research in Science Education, vol. 19, 291 – 299.

Webb, M. J. (1985). *Analogies and their limitations.* School Science and Mathematics, vol. 85, 645 – 650.

Yin, R. K. (1989). *Case study research: design and methods* (2nd ed.). Beverly Hills, CA: Sage Publications.

Zeitoun, H. H. (1984). *Teaching scientific analogies: A proposed model.* Research in Science and Technological Education, vol. 2 (2), 107 – 125.

Textbooks analyzed

Ainley, D., Lazonby, J. N. & Masson, A. J. (1981). *Chemistry in today's world.* London: Bell & Hyman Limited.

Boden, A. (1986). *Chemtext.* Marrickville, NSW: Science Press.

Bucat, R. J. (Ed.). (1983). *Elements of chemistry* (plain vol. 1,). Canberra: Australian Academy of Science.

Bucat, R. J. (Ed.). (1984). *Elements of chemistry* (plain vol. 2,). Canberra: Australian Academy of Science.

Elvins, C., Jones, D., Lukins, N., Miskin, J., Ross, R. & Sanders, R. (1990). *Chemistry in context: Part one.* Melbourne: Heinemann Educational Australia.

Garnett, P. J. (Ed.) (1985). *Foundations of chemistry*. Melbourne: Longman Cheshire Pty Limited.

Hunter, R. J., Simpson, P. G. & Stranks, D. R. (1981). *Chemical science*. Marrickville, NSW: Science Press.

James, M., Derbogosian, M., Bowen, S. & Auteri, S. (1991). *Chemical connections*. Melbourne: The Jacaranda Press.

Lewis, P. & Slade, R. (1981). *A guide to H.S.C. chemistry*. Melbourne: Longman Cheshire Pty Limited.

McTigue, P. T. (Ed.) (1979). *Chemistry — key to the earth*. Melbourne: Melbourne University Press.

COMMENTARY OF THE DISCUSSANT

Christian Otzen, Pädagogische Hochschule Kiel, Germany

For the organisation of learning processes the use of analogies increasingly becomes the focus of attention by pedagogues and didactic specialists. As results of the research in psychology of learning seem to verify the benefits of correctly used analogies, this is now a topic within the didactic methodical discussion in chemistry. I do appreciate this because students of chemistry rather frequently have great difficulties in understanding and therefore this subject is quite unpopular — so it urgently needs new impulses.

In view of investigation results about systematically use of analogies in chemistry education we might be quite optimistic about their benefit. Inter alia you can draw the following conclusions from these research results:

- By using analogies new possibilities can be presented for the discovering of solving strategies.
- Analogies are often helpful to develop a motivational effect on students' learning attitude.
- Analogies can provide bridges between abstract and concrete concepts and vice versa.
- Analogies can help to assimilate new material with prior knowledge.
- Analogies can be suitable to create a link between scientific matters and students' real world experiences.

Any of these conclusions gives most interesting perspectives for the didactic and methods of chemistry education.

By investigating textbooks David Treagust and R. B. Thiele found out that analogies are extensively employed to explain matters and concepts in chemistry. In my estimation this does not hold to such an extent for German textbooks. I just want to quote one analogy which is quite popular here to explain the phenomenon of chemical activation energy: by using energy a car is pushed onto the top of a hill so that finally it can get down easily on its own.

Far more often analogies are used spontaneously by teachers and students during classroom instructions. In this case David Treagust's results correspond to my observations as a teacher of chemistry for many years in Schleswig–Holstein/Germany and as well to my experiences in teacher training. At first sight the use of analogies seems to provide some positive effects on learning processes and seems to be specially beneficial for chemistry education, but I doubt that the use of analogies in the practice of school education might stand up to a critical review.

I totally agree with David Treagust when he says that there is "need for research to investigate for whom and under what conditions analogies are most beneficial."

From the results of David Treagust's interviews with teachers you can get the impression that the selection of spontaneously employed analogies is more due to the principle of trial and error or individual intuition but less to careful planning. This is not surprising as there are not any scientifically determined criteria for the selection of analogies.

My critique is based on the following reflections:

(1) It is uncertain that if a teacher decides to use an analogy this is because of the students having comprehension problems or does the teacher lack competence in explaining a special matter. In that case the use of an analogy would just represent the teacher's pedagogical difficulty but would not primary be intended to aid the students' understanding.

(2) To be effective, analogies have to correspond to students' real world experiences, to their special interests and to their prior knowledge. But the present school reality — at least in Germany — only gives a little opportunity to the teachers to know their students sufficiently well so that they could select analogies according to the requested sense. Analogies which are just based on the teacher's world of thoughts may not be the right ones for the students' cognitive structures.

(3) An efficient use of analogies demands that the attributes of the analogy and the target content do not differ too much. So in a concrete teaching situation the teacher has to find an analogy which at the same time has to correspond to students' understanding and correctly to the target matter. Spontaneously this problem can certainly be solved in only a few cases.

(4) For the students analogies often are not really a simplification of a difficult material. For a cognitive comprehension they have to abstract from a concrete situation. So if the students already have difficulties to understand the basic problem they will of course have difficulties in doing the abstraction.

(5) Schoolclasses normally consist of students with different learning backgrounds. So there is the danger that only some of the students can profit beneficially from an analogy whereas for others analogies possibly just cause new difficulties or even lead to false results.

On the whole, analogies which are useless, incorrectly selected or not employed properly may rather damage the process of learning and understanding than aiding it beneficially. They might create new problems before the target concept is understood, they can provide false interpretations of a matter or may just confuse the students. So the effect of a false use of analogies may be contraproductive to a learning process for a long time.

In this sense E. Kircher partly regards the present way of employing analogies in physics education to be a waste of time. For chemistry education I would underline this statement.

Many of the stated aspects had already been mentioned in David Treagust's report. It is important to me that because of the stated advantages of analogies in chemistry education to specially emphasise its dangers. We have to be aware of the dangers so that we can profit from the in fact existing benefits.

To sum up:

(1) Analogies have to be an aid for understanding and may not be used to cover a teacher's pedagogical inability.

(2) The right selection of an analogy is a difficult task, and it needs a careful analysis of the conditions under which should be learned.

(3) Analogies have to be critically employed and their effects have to be controlled.

(4) It still needs intensive investigation to evaluate possibilities and limits of the use of analogies.

References

Dörner, D. (1974). *Die kognitive Organisation beim Problemlösen*. Bern: Huber.

Druit, R. (1991). *On the Role of Analogies and Metaphors in Learning Science*. Science Education, vol. 75 (6), 649 – 672.

Kircher, E. (1989). *Analogien im Physikunterricht*. In: Schneider, W., Wege in die Physikdidaktik. Erlangen: Palm und Enke.

Kluwe, R. (1979). *Wissen und Denken*. Stuttgart: Kohlhammer.

Otzen, Ch. (1990). *Chemieunterricht in der Bundesrepublik Deutschland — Anspruch und Wirklichkeit*. In: Akademie der Pädagogischen Wissenschaften der DDR (Ed.), Bewährenswertes und Erneuerungsbedürftiges im Mathematischen und Naturwissenschaftlichen Unterricht, Berlin.

SUMMARY OF THE PLENARY DISCUSSION

Jutta Theißen, Universität Dortmund, Germany

To show the difference between a model and an analogy the term "model" was defined. A model attempts to be an analogy, it has analogical features. It represents a phenomenon and models the target. For example, the model of moving balls is an analog to moving gas molecules. It is very important to point out the parts that cannot be mapped, i. e. where the model breaks down.

Analogies are available at school, in textbooks many possibilities can be found. However, the teachers may not be trained to teach them. Hopefully, they can integrate analogies in their teaching, but unfortunately, a lot of teachers do not have a good repertoire, and, moreover, do not map and show where the analogy breaks down.

The lack of time was another item of discussion. Not having enough time was considered

ERIC

as being the big dilemma of science education. A teacher's efficiency often is determined by how many of his/her students pass the exams. The teacher has to fulfill the objectives of the syllabus, to cover the content as soon as possible in order to enable the students to pass the exams. However, teaching analogies does not take much time, so that though being pressed for time the teacher can afford to teach analogies.

It was pointed out that the target topic should always be more important than the analogy. When using analogies for teaching teachers present analogies, show the similarities, and then break down the analogies. Exam questions do not refer to analogies.

Analogies are never useful for all students. Some of them may find it a painful exercise to learn the analogy and understand the chemical target topic, but a lot of students do get along with it and learn by analogies.

One has to be very careful when using analogies, and should only choose useful analogies. If models or analogies do not fit very well they may not only not help the students, but even confirm or create misconceptions. To avoid this, students' theoretical framework should be examined to get to know their misconceptions. Furthermore, the teacher should show similarities and dissimilarities, and break down the analogy in the end. The following model may help teachers to teach analogies:

(1) Introduce the target concept to be learned.
(2) Cue the students' memory of the analogous situation.
(3) Identify the relevant features of the target concept and the analog.
(4) Map out the similarities between target concept and analog.
(5) Draw conclusions about target concept.
(6) Indicate where the analog breaks down.

It also would be an interesting idea to have the students develop their own analogies, but in order to be able to develop an analogy you have to understand the target first.

ERIC

List of authors

Dale Baker, Arizona State University, College of Education, Division of Curriculum and Instruction – Elementary Education, Tempe, Arizona 85287 – 0911, United States

Robert Fairbrother, University of London, King's College, Conwall House Annex, Waterloo Road, SEI 87 X London, United Kingdom

Myra J. Halpin, The North Carolina School of Science and Mathematics, 1219 Broad Street, Durham, North Carolina 27715, United States

Hanno van Keulen, University of Utrecht, Centre for Science and Mathematics Education, Princetonplein 1, 3508 Utrecht, The Netherlands

Gerhard Meyendorf, Dürerstraße 4, O–4370 Köthen, Germany

Michael D. Piburn, Arizona State University, College of Education, Tempe, Arizona 85287 – 0211, United States

Iris Pigeot–Kübler, Universität Dortmund, Fachbereich Statistik, Emil Figge Straße 70, 4600 Dortmund–Eichlinghofen, Germany

Peter van Roon, University of Utrecht, Centre for Science and Mathematics Education, Princetonplein 1, 3508 Utrecht, The Netherlands

Elisabeth Schach, Universität Dortmund, Hochschulrechenzentrum, August–Schmidt–Straße 12, 4600 Dortmund–Eichlinghofen, Germany

Hans–Jürgen Schmidt, Universität Dortmund, Fachberech Chemie, Otto–Hahn–Straße 6, 4600 Dortmund–Eichlinghofen, Germany

Amos A. M. Shaibu, Ahmadu Bello University, Department of Education, Science Education Section, Zaira, Nigeria

Rodney B. Thiele, Curtin University of Technology, Science and Mathematics Education Centre, Perth 6001, Western Australia

David F.Treagust, Curtin University of Technology, Science and Mathematics Education Centre, Perth 6001, Western Australia

ERIC

Every other year science education researchers and statisticians meet at the Dortmund Summer Symposium to present papers and discuss their studies. The 1992 Proceedings contain all the papers as well as the summarises of the discussions. The topics brought forward are of great interest to all involved in basic research in science education

I C A S E

ERIC Full Text Provided by ERIC

BEST COPY AVAILABLE