DOCUMENT RESUME

ED 481 658	TM 035 358
AUTHOR	Popp, Sharon Osborn; Behrens, John T.; Ryan, Joseph M.; Hess, Robert K.
TITLE	Applying the Rasch Model for Ordered Categories To Assess the Relationship between Response Choice Content and Category Threshold Disorder.
PUB DATE	2000-04-00
NOTE	16p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 2000).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Classification; *High School Students; High Schools; Item Response Theory; *Responses; Sciences; *Student Attitudes
IDENTIFIERS	*Rasch Model; *Thresholds

ABSTRACT

The Rasch model for ordered categories was applied to responses on a science attitude survey that uses a combined semantic differential and Likert-type scale format. Data were drawn from the Views about Science Survey for 1,300 high school students. Examination of category response function graphs and threshold estimates allowed classification of items into three patterns of threshold disorder. The three patterns provided insight into the degree of content polarization between endpoint response choices (e.g., items with highly polarized response choice content produce responses toward the extremes and have disordered thresholds or compressed threshold range among the central categories). The patterns were used to direct modification of the response format with respect to number of choices and extremity in endpoint wording. (Contains 2 figures and 17 references.) (Author/SLD)

Applying the Rasch Model for Ordered Categories to Assess the Relationship Between Response Choice Content and Category Threshold Disorder

Sharon Osborn Popp and John T. Behrens

Arizona State University

Joseph M. Ryan and Robert K. Hess

Arizona State University West

A paper presented at the Annual Meeting of the American Educational Research Association, New Orleans. April, 2000.

Supported in part by a grant from the National Science Foundation

Abstract

The Rasch model for ordered categories was applied to responses on a science attitude survey that employs a combined semantic differential and Likert-type scale format. Examination of category response function graphs and threshold estimates allowed classification of items into three patterns of threshold disorder. The three patterns provided insight into the degree of content polarization between endpoint response choices (e.g., items with highly polarized response choice content produce responses toward the extremes and have disordered thresholds or compressed threshold range among the central categories). The patterns were used to direct modification of the response format with respect to number of choices and extremity in endpoint wording.

Using the Rasch Model for Ordered Categories to Assess the Relationship Between Response Choice Content and Category Threshold Disorder

Patterns of category threshold disorder were observed to be related to the degree of polarization in response choice content on items from a survey instrument used to assess the relationship between student views and science achievement. Survey items with high response choice content polarization produce responses toward the extremes and have disordered thresholds or compressed threshold range among the central categories. Items with low content polarization produce less response variability and show threshold disorder or compression within the outer response catgeories. The relationship observed between the types of category threshold disorder and response choice content suggest that modifications to the number of response categories and wording used in endpoint labels may be essential to successful implementation of the response format employed by the instrument.

The instrument, Views About Sciences Survey (VASS), employs a novel response format called a Contrasting Alternatives Design (CAD) (Halloun and Hestenes, 1996). Under the CAD's associated scoring system, analyses based on classical test theory have failed to provide strong evidence to support the validity of VASS or identify the source of problems. This study (1) reviews the obstacles to scoring VASS responses under the current response category structure, (2) reports the results of analyses using the Rasch model for ordered categories (RMOC) (Andrich, 1978; Masters, 1982) with this instrument, and (3) discusses the relationship between content of item response choices and category disorder, and implications for item construction revealed by this analysis.

VASS and the Rationale for the CAD Response Format

VASS consists of a series of 30 items developed to characterize student views about knowing and learning science (Halloun and Hestenes, 1996; Halloun, 1997). The survey was developed for use in a National Science Foundation funded physics education reform project, and has been used in universities and colleges to assess the effects of implementing reform methods of science instruction. Halloun and Hestenes had found that available measures used to assess student views about science were problematic in terms of reliability and validity (Halloun, 1994; Munby, 1983; Rennie and Parker, 1987). This conclusion echoes concerns expressed by science education researchers regarding the need for improved attitude assessment instruments (Haladyna, Olsen and Shaughnessy, 1983; Krynowsky, 1988; Schibeci, 1984; Willson, 1983).

Early, constructed-response versions of VASS were piloted, but interviews held with students often yielded information contradictory to the student's written responses. Halloun and Hestenes (in press) give an example of one student's response to an essay question, where students had been asked to state the first thing they do in solving physics problems. The student responded that he starts by looking for the appropriate formula. However, during an interview, the student revealed that he actually starts to solve a physics problem by drawing diagrams, but had not thought that this was worth mentioning in his written response. They then discuss the problems associated with transforming such a question into a traditional survey format such as a Likert-type scale or multiple choice. VASS's Contrasting Alternatives Design (CAD) was developed to assess where an individual falls along a continuum between two different perspectives (that may not be completely at odds with each other) in order to assess the degree to which students differ from experts in their views about knowing and learning science and to address the shortcomings encountered with available instrument formats.

Traditional formats may not be well-equipped to capture the gradations between contrasting views which are not necessarily diametrically opposite. Gardner (1987) presents evidence regarding conventional use of Likert-type and semantic differential scales in measuring attitude toward science and shows that "favorable and unfavorable statements are not necessarily bipolar opposites" (p. 245). He called for new psychometric approaches, such as analyzing conventional instruments differently, employing new scales that separate concepts, or the measurement of ambivalence directly. The CAD format of VASS could be characterized as an attempt to realize the direct approach. Many of the pairs of contrasting views put forth in VASS items may elicit some degree of agreement toward each option presented. The degree of imbalance between the alternatives is what the CAD format is intended to assess.

Halloun (1997) outlines the views that VASS is intended to measure within the context of several dimensions (Learnability; Reflective Thinking; Personal Relevance; Structure; Methodology; Validity). He contrasts the views usually held by scientists and educators (reflecting scientific realism and critical learning) and views often held by the lay community and many students (naïve realism and passive learning). These contrasting views may be better described as distant points along a continuum of perspective, rather than opposing ends of a strictly bipolar evaluative dimension. For example, the Learnability dimension of VASS includes items designed to assess whether "[a]chievement depends more on *personal effort* - than on the influence of teacher or textbook" (Halloun, 1997). The expert view on this and related items would certainly emphasize personal effort, but may not completely reject the influence of teacher and textbook. The lack of a clear-cut "agree/disagree"or "either/or" in the constructs to be measured needed to be considered in the development of an instrument. Halloun and Hestenes developed a response scale that differentially weights the expert-emphasized view with the naïve view in the Contrasting Alternatives Design (CAD) format.

A distinctive feature of VASS's CAD format is that it contains elements of both a semantic differential scale and a Likert-type scale format. CAD items consist of an incomplete statement followed by two contrasting alternatives that may complete the statement. Figure 1 gives an example item from VASS.

After the teacher solves a physics problem for which I got a wrong solution on my own:

a) I discard my solution and learn the one presented by the teacher.
b) I try to figure out how the teacher's solution differs from mine.

Answer Options :

1	2	3	4	5	6	7	8
Only a, Never b	Mostly a, Rarely b	More a than b	Equally a and b	More b than a	Mostly b, Rarely a	Only b, Never a	Neither a nor b

Figure 1. An item from VASS

The alternatives for each item represent an "expert" view, typically held by professors and teachers of physics, or a naïve or "folk" view, typically held by students or lay persons. Respondents are asked to choose from a continuum of seven possible responses that are ordered in a weighted manner representing degree of preference for one alternative compared to the other, or may choose an eighth response if they do not agree with either alternative to any degree.

VASS items differ in the degree of content polarization between the contrasting alternatives. Some pairs of contrasting alternatives are nearly or literally mutually exclusive (e.g., "If I had a choice:..." a: "I would never take any physics course" vs. b: "I would still take physics for my own benefit"), while others are more compatible (e.g., "For me, doing well in a physics course depends on:..." a) "how much effort I put into studying" vs. b) "how well the teacher explains things in class").

A conventional summated ratings approach has not provided meaningful scoring of responses to the CAD format of VASS. Previous samples of VASS responses yielded rather low estimates of internal consistency (Cronbach's alpha values of .59 - .64) and low item-total correlations associated with several items. However, response patterns (i.e., proportions of endorsement to each response category) were very consistent across samples and clustering. The authors of the instrument implemented a recoding of responses (by collapsing across categories) which showed relationships between performance on VASS and achievement criteria. However, there was little empirical justification for the collapsing procedures that varied by item. Variable rating scale widths within each item and among items, as well as irregular ordering of responses on some items, were suspected to be limiting the ability of VASS to represent the hypothesized measurement construct of folk-expert view.

Scoring Challenges and Content Polarization Issues

The major challenges to summation of ratings in scoring VASS's CAD format are (a) variable rating scale widths within items, (b) variable rating scale widths among items, and (c) irregular ordering of response continuum on some items. Variable rating scale widths within and among items are cause for concern regarding the conventional use of summated ratings with any measurement instrument. Thorndike (1904) refers to inequalities of units as one of the "special difficulties" in social science measurement. He illustrated the problem in his criticism of a spelling test. "Thorndike argued that the correct spelling of an easy word versus a hard word did not reflect equal amounts of spelling ability" (Engelhard, 1992, p. 284). Irregular ordering of the response categories is also problematic for the use of conventional summated ratings. For the summation of ratings to be meaningful, response data must possess the property of conjoint transitivity, i.e., a hierarchical ordering of responses that correspond to the increasing degree of the underlying measurement construct.

Variable rating scale widths within items may be strongly influenced by the degree to which the contrasting alternatives provided in particular VASS items are bipolar. Items where the alternatives are less polarized, and presumably might elicit less extreme responses, predictably result in a more centrally crowded distribution of the responses. Wyatt and Meyers (1987) found that the degree to which the respondents were prone to polarize their responses due to strongly held opinions, affected response variability depending on whether scale endpoints were more or less "nearly absolute". For respondents holding stronger opinions, "a more nearly absolute scale might be used to draw responses toward the middle of the scale" (Wyatt and Meyers, 1987, p. 33), while a less absolute scale could be used to encourage response variability among respondents that do not hold extreme views. Lam and Stevens (1994) also looked at the impact of rating scale design on item variability and found differences depending on degree of content polarization, the degree to which the endpoint labels were absolute, and intensity of item wording.

The degree of bipolarity between contrasting alternatives affects the meaning of the response to each item, relative to responses on other items, not just other potential responses within the item. This results in variable rating scale widths among items. A response of 5 falls within the range of expert-level response on some low-bipolarity items. A response of 5 falls within the range of a mixed-view response on other items with more polarized alternatives. While summation of responses may yield a rank ordering of respondents by raw score, the ordering may not correspond directly to the degree of the theoretical expert-view latent trait held by respondents.

The irregular ordering of within-item responses on some VASS items is a serious threat to the use of summated ratings. Response categories must have a hierarchical relation for the sum of observed responses to reflect the underlying construct in a meaningful way. It follows that a person with a greater degree of overall expert view toward science should have a higher probability of achieving a greater score on a particular VASS item (i.e., higher rating). However, results of early content validation confirmed that the underlying construct on certain VASS items does not support this assumption. The expert responses of university and high school instructors revealed that some items elicited the majority of actual expert responses on the penultimate expert-pole response, not on the most extreme expert-pole response (Halloun and Hestenes, 1996).

Method

Data

The data used in this study were collected as part of a National Science Foundation-sponsored high school physics education reform project, based on the

Modeling Method of physics instruction (Wells, Hestenes, and Swackhamer; 1995). The Views About Sciences Survey (VASS) is among several instruments used in the assessment and evaluation of the project. Data was collected from 1293 students of 45 high school teachers from 13 states during the 1996-1997 academic year. Data were then collected from 2123 students of 51 high school teachers from 26 states during the 1997-1998 academic year. The VASS was administered by the teachers, during regular classroom hours, and within the first weeks of instruction. A random sub-sample of 1300 subjects was drawn from the 1997-1998 sample to facilitate comparison of sample-size sensitive fit values between the two sets of responses.

Procedure

The RMOC was applied twice to the 1996-1997 sample. The first analysis examined the performance of the original seven response categories per item. Item fit statistics were examined and compared to results from classical item analyses.Threshold parameter estimates and category response function (CRF) graphs were examined and provided diagnostic information regarding disordered categories on most items. Three category-collapsing strategies were developed based on the observed patterns of threshold disorder. The second analysis of the 1996-1997 sample tested hypothesized modifications to the number of response categories for particular sets of items by applying the three collapsing strategies. The same collapsing strategies were then applied to the 1997-1998 sample to assess the performance of the recommendations under cross-validation.

Results

The seven-category model produced disordered threshold parameter estimates for 24 of the 30 VASS items; the remaining six items showed some degree of compression among the inner category response thresholds. Patterns of threshold parameter estimates could be classified into three types: 1. very compressed range among all five of the inner categories threshold values (often overlapping and reversed), 2. compression of the most central threshold values without affecting outer categories, and 3. reversal or compression in the outer threshold values without affecting the most central categories. Figure 2 below provides an example category response function graph for each pattern type.

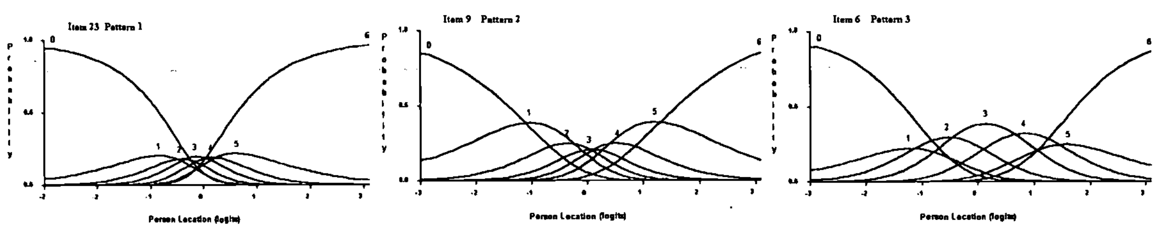

Figure 2. Examples of three common CRF graph types for 1996-1997 VASS items

Three category-collapsing strategies were developed to address the three patterns of disorder. The items that produced threshold estimate patterns like the first pattern have the most bipolar content in their contrasting alternatives; these categories were collapsed from seven to two categories. The items that produced patterns like the second pattern have contrasting alternatives with a high degree of content polarization; however, the associated alternatives are not mutually exclusive. These categories were collapsed from seven to five categories, with inner categories being collapsed. The items that produced patterns like the third pattern have contrasting alternatives with a low degree of content polarization; these categories were collapsed from seven to five categories, with the outer categories being collapsed. The collapsing strategy associated with the second pattern was applied to the six items that showed inner threshold compression, but did not produce disordered thresholds.

The analysis of the 1996-1997 sample with category-collapsing produced an item-trait-interaction total chi-square value of 326.776 ($\underline{N}$ = 1293; $\underline{df}$ = 58; $\underline{p}$ < .001 compared to 624.802 ($\underline{N}$ = 1293; $\underline{df}$ = 58; $\underline{p}$ < .001) for the seven-category model. No disorder was observed among the threshold parameter estimates. Application of the category collapsing strategies on the 1997-1998 sample produced three items with disordered threshold estimates. However, the values of these threshold estimates were not significantly different from the values produced with the 1996-1997 sample.

Discussion

Overall, CRF graph pattern types appear to be related to the degree of content polarization between the contrasting alternatives on sets of VASS items. This is consistent with research that shows that the degree to which rating scale endpoints are absolute interacts with the degree of content polarization in item response choices. For example, on items which show the third pattern (i.e., items with low content polarization), the absolute endpoints of VASS response choices appear to be reducing response variability by drawing respondents toward middle categories. These items evoke less extreme opinions with respect to the relatively more compatible contrasting alternatives.

The patterns of category disorder that led to the recommended revisions provided insight into the wording of VASS item response alternatives. Variations on the number and labeling of response categories, depending on the content of the response alternatives, may be needed for the successful implementation of the CAD format. Results suggest that modifying the number of response categories of VASS items to prevent category disorder may allow performance on the VASS to better reflect the intended measurement construct. More "expert view" responses on each item will more directly correspond to an overall more "expert view". The degree of content polarization on the three type of items can also provide direction regarding modification of endpoint labels, in addition to number of categories.

Considering the RMOC results it is recommended that Pattern 1 items have a dichotomous response format, consisting of only the two contrasting alternatives. Pattern 2 and 3 items would have five as the optimal number of categories, with two different types of response category labeling. Pattern 2 items would keep extreme wording (i.e. Only [a], Never [b]) in endpoint labeling to increase probability of response in the central categories. The near-extreme responses could have the current, more central response alternatives (i.e. More [a] than [b]). Pattern 3 items would have less extreme wording in the endpoint labels. Endpoint labels could be the same as the current near-extremes (i.e. Mostly [a], Rarely [b]).

Clearly, empirical validation of the all recommended changes in number of response categories and endpoint labels is recommended for a modified version of VASS. The success of the category-collapsing strategies offers hope that the hypothesized underlying measurement construct of folk-expert view can be better measured and understood with a modified version of the instrument. The RMOC offered an effective approach for quantifying the problems underlying the complex data, as well as a means to refine the content of the items and understand the implications of the wording of item response choices.

References

Andrich, D. (1978). A rating formulation for ordered categories. *Psychometrika, 43*, 357-374.

Engelhard, G. (1992). Historical views of invariance: Evidence from the measurement theories of Thorndike, Thurstone, and Rasch. *Educational and Psychological Measurement, 52*, 275-291.

Gardner, P. L. (1987). Measuring ambivalence to science. *Journal of Research in Science Teaching, 24*, 241-247.

Haladyna, T., Olsen, R., & Shaughnessy, J. (1983). Correlates of class attitude toward science. *Journal of Research in Science Teaching, 20*, 311-324.

Halloun, I. (1994). Assessing student views about physics. How adequate are available instruments? Invited paper presented at the Annual Meeting of the American Association of Physics Teachers, Notre Dame, IN.

Halloun, I. (1997). Views about science and physics achievement: The VASS story. In E. F. Redish and J. S. Rigden (Eds). *The Changing Role of Physics Departments in Modern Universities, Proceedings of the International Conference on Undergraduate Physics Education, College Park, MD, 1996*. (pp. 605-613). Woodbury, NY: American Institute of Physics.

Halloun, I., & Hestenes, D. (1996). Views about sciences survey. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, St. Louis, MO.

Halloun, I., & Hestenes, D. (in press). Interpreting VASS dimensions and profiles. Science and Education.

Krynowsky, B. A. (1988). Problems in assessing student attitude in science education: A partial solution. *Science Education, 72*, 575-584.

Lam, T. C. M., & Stevens, J. J. (1994). Effects of content polarization, item wording, and rating scale width on rating response. *Applied Measurement in Education, 7*,141-158.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149-174.

Munby, H. (1983). Thirty studies involving the "Scientific Attitude Inventory": What confidence can we have in this instrument? *Journal of Research in Science Teaching, 20*, 141-162.

Rennie, L. J., & Parker, L. H. (1987). Scale dimensionality and population heterogeneity: Potential problems in the interpretation of attitude data. *Journal of Research in Science Teaching, 24*, 567-577.

Schibeci, R. A. (1984). Attitudes to science: An update. *Studies in Science Education, 11*, 26-59.

Wells, M., Hestenes, D., & Swackhamer, G. (1995). A modeling method for high school physics instruction. *American Journal of Physics, 63*, 606-619.

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, 20*, 839-850.

Wyatt, R.C., & Meyers, L. S. (1987). Psychometric properties of four 5-point Likert-type response scales. *Educational and Psychological Measurement, 47*, 27-35.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

TM035358

I. DOCUMENT IDENTIFICATION:

Title: Applying the Rasch model for ordered categories to assess the relationship between response choice content and category threshold disorder	
Author(s): Osborn Popp, S. E.; Behrens, J. T.; Ryan, J. M.; & Hess, R. K.	
Corporate Source:	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: [signature]	Printed Name/Position/Title: Sharon E. Osborn Popp, Faculty Associate/ Research Associate, Arizona State University	
Organization/Address: 4531 W. Toldeo Street Chandler, AZ 85226	Telephone: (480) 705-0256	FAX:
	E-Mail Address: osbornpo@asu.edu	Date: 11/07/2003

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: info@ericfac.piccard.csc.com
WWW: http://ericfacility.org

EFF-088 (Rev. 2/2003)