NEW ELECTRONIC TECHNOLOGIES FOR FACILITATING DIFFERENTIATED INSTRUCTION

By

KATHLEEN SCALISE *

* *Assistant Professor, Educational Leadership, University of Oregon, Eugene, USA*

ABSTRACT

With electronic technologies, differentiated instruction has the same meaning as in traditional instruction, but different tools are available for teachers to help students learn. Electronic technologies for differentiated instruction can add powerful new types of media inclusion, levels of interactivity, and response actions. This rapidly emerging approach to differentiated instruction also can enhance ability to collect data on the fly and to deliver custom content. This paper considers some self-selection strategies for differentiated instruction online, and presents the results of one study to explore student use of self-selection.

Keywords: Differentiated Instruction, E-learning, Computer Adaptive, Item Response Models, Iota Model, Modify Option, Computer-based Testing, Science Education.

INTRODUCTION

Differentiated instruction is an approach to teaching that acknowledges people have multiple paths for learning and for making sense of ideas (Hall, 2002; Reis, et al., 1988; Sizer, 2001; Tomlinson, 2001; Tomlinson & Allan, 2000; Tomlinson & McTighe, 2006; Willis & Mann, 2000). With the use of electronic technologies, differentiated instruction is beginning to play out in some new forms (Scalise, 2005; Taylor, 2002; Trivantis, 2005; Turker, Görgün, & Conlan, 2006). These include new media inclusion for differentiation, levels of interactivity, response actions, and enhanced ability to collect data on the fly and to deliver custom content (Bennett, 2000; C. G. Parshall, Spray, Kalohn, & Davey, 2002; C. G. Parshall, Stewart, R., Ritter, J., 1996).

Here the author present some results from a study within a differentiated e-learning product for which students are allowed to modify selected-response question answers, if they answer they would like is not listed. The study was undertaken to see if self-selection of different answer choices would be favored by students in differentiated e-learning products online, and if it would produce new differentiation evidence.

Understanding How Assessments in Differentiated E-Learning Are Used

Often in differentiated e-learning products, assessment approaches are being used formatively, or in other words to guide instruction during the process of learning, Here differentiating may be done by the challenge level, types of formats, representations and feedback (Black, Harrison, Lee, Marshall, & Wiliam, 2002; Black & Wiliam, 1998). Assessments are used to drive what content is offered to each student usually with the intention of making a difference in how or how much the child or adult learns.

Information can also be used summatively, or in order to make a judgment about student learning. This includes appropriate course placement or who gets access to what educational opportunities (Resnick & Resnick, 1992). Feed forward to teachers by systems that collect and report student information can also influence teacher expectations of students. Taken all together, the potential of differentiation to affect student learning can be great (Tomlinson & McTighe, 2006). In the e-learning context, it also becomes faster and easier to do for some types of differentiation, so it is important that differentiation is well done, just as is true in the classroom-based context.

Distinctions important for instructional leaders to understand

When instruction is differentiated in the classroom, it is often clear that multiple approaches are spiraled into the curriculum. For instance, experiences repeat in different forms or students are grouped and regrouped for course placement and learning activities.

With electronic technologies, however, it can be much less apparent that differentiation is taking place. If one learner is given something different on the computer from some other learner, either locally or at a remote site, it can be hard to tell since the two learners aren't looking at the same screens. Typically there is no basis for comparison. The learner may not even realize that had he/she interacted differently with the computer, it would have interacted differently with them. Also, unless disclosed, we don't necessarily know what e-interfaces are gleaning about a learner or the purposes to which the inferences are being put (Nielsen, 1998).

Furthermore, differentiation in e-learning products can have a different intent from classroom-based approaches. In traditional classroom-based approaches, some researchers argue that the result of differentiated instruction should not be different learning outcomes but rather different ways to access the same learning outcomes (Tomlinson & McTighe, 2006). The argument often is that the strongest classroom-based differentiation ensures all students work with the essential understandings for a segment of learning, thus stabilizing the most substantial learning goals. E-learning products are often designed to stretch the individual student's opportunity to learn. The intent can be to tap areas were interest and engagement are strong, or to give the learner choice among objectives.

Results of Self-Selection in Assessment

While many differentiation strategies use model-based assessments such as item response models and Bayesian networks to generate evidence for differentiation, a different approach relies on self-determination. Multiple possible paths of learning are made available through the electronic technology. Students select their personal choices as they go. This can consist of simply selecting the order of completion among a fixed menu of learning activities or course modules. More flexibility and choice comes about when students get to select from among a range of different activities, leaving some out and doing others.

Self differentiation is a very common type of differentiation seen in e-learning content. E-learning environments are often built on a “hyperlink” paradigm, such as seen on the Web, where links in the content can be self-selected for additional information or learning opportunities.

An example of self differentiation in an adult learning context are the e-learning products of the Collaborative IRB Training Initiative (CITI). CITI offers instructional modules for understanding protection of human research subjects. Each module is focused on a different aspect of bio-ethics and human subjects research. Here the self-selection differentiation is by course module. For CITI certification, completion of a uniform set of course modules is required. But an additional set of self-selected course modules are also required. The group of self-selected modules covers different topics, such as human subjects protections for special needs populations or for research in schools. The learner is able to select the topics they feel are most appropriate to their learning needs.

In regards to assessment, questions, tasks, activities and other methods of eliciting student responses are often called items in the assessment process. In the computer-based platform, the authors argue that almost any type of interaction with a user can be considered as an assessment item. Note that a working definition they have proposed for an assessment item (when speaking in reference to CBT) is any designed interaction with a respondent from which data is collected with the intent of making an inference about the respondent.

Specifically the authors will address results of a “modify” option included in some of the questions. With the "modify" option, students could choose to select one of the selected response answers and then modify it to put the answer in their own words. This item design innovation was met with much enthusiasm by both subject matter experts in the area of the examination and by measurement review panels, as a technique of exploring examine desire for further constructed response opportunities within some of the innovative formats.

In introducing the modify option, they did not know how many students would choose to rewrite the answers, nor did they have a good sense of how the option would be used. Pilot trials suggested that the higher the expertise of the respondent, the more likely they were to choose to restate answers in their own words, and that students who felt they were busy and had limited time available to complete the assignment were less likely to rewrite answers.

Sample

In this paper, the authors model testlet data from the UC Berkeley "Smart Homework" implementation of ChemQuery, an NSF-funded project, in which one component consisted of data driven content to individually tailor "smart" homework sets in high school and university level chemistry. The adaptivity approach of the homework sets is called BEAR CAT, since it draws on a CAT approach to the BEAR Assessment System (Wilson, 2005) for the specification of properties, or variables, of interest to measure. The BEAR CAT approach is a multistage CAT, in which sequential or preplanned pathways are adaptively presented to students within the testlets based on student responses, and updating of theta and standard CAT algorithms can be used between the testlets. In order to investigate a variety of instructional design approaches within the BEAR CAT bundles, three testlet designs were invented and used across content areas. The three bundle designs differed in the target level of the opening question, the number of allowed paths to same score, and the range of item formats employed within the testlet.

Data were collected from 521 students involved in the BEAR CAT study. To administer the adaptive testlets, the BEAR CAT smart homework sets were deployed through the Homework Tool capabilities of the Distributed Learning Workshop Learning Management System (Gifford, 2001), with the Homework Tool modified to accommodate the adaptive instructional flow.

Psychometrically, context effects and inter-item dependence are a threat to testlets, and need to be modeled by correct statistical models. Important sources for formal modeling options for testlets include Li, Bolt and Fu (2006), Wilson and Adams (1995) and Wainer and Kiely (1987). In testlet structures, clustered items usually are linked by attributes such as common stimulus material and common item stem, structure and content (Wilson & Adams, 1995). This suggests the usual assumption of conditional or local independence between items necessary for item response modeling is not met within the testlet. Local independence in item response models "means that the response to any item is unrelated to any other item when trait level," or student performance level, "is controlled" (Embretson & Reise, 2000). The local independence assumption is commonly stated as shown below in Equation 1:

$$P(X_{is}=1 \mid X_{js},\xi_i,\theta_s)=P(X_{is}=1 \mid \xi_i,\theta_s) \qquad (1)$$

where X_{is} represents the score of student s on item i, X_{js} represents the score of the same student on another item j, $_{i}$ represents a vector of item parameters for item i, and θ_s represents the person performance ability of student s. One approach for addressing such within-bundle dependence is to treat each bundle of dependent items as a single item, awarding degrees of partial credit over the testlet depending on level of overall performance indicated by the series of responses, which can be called the bundle response vector (Wilson & Adams, 1995).

Testlets previously have been psychometrically modeled in a variety of ways, most usually with some version of a partial credit model. The partial credit model is the more general of two polytomous Rasch models (Wright & Masters, 1982) commonly expressed according to Equation 2:

$$P(X_{is}=x \mid \theta_s)=\frac{\exp\sum_{j=0}^{x}(\theta_s-\delta_{ij})}{\sum_{r=0}^{m_i}\exp\sum_{j=0}^{r}(\theta_s-\delta_{ij})} \qquad (2)$$

for item i scored x=0,…,m_i, where X_{is} is the score of student s on item i, x represents a given score level, θ_s represents the parameter associated with the person performance ability of student s, r in the denominator represents a summation of terms over the steps, and δ_{ij} represents the parameter associated with the difficulty of the jth step of item i.

Another modeling approach developed by Howard

Wainer is testlet response theory (H. Wainer, et al., 2006), with a 3PL multi-faceted model "testlet effect" which is a special student ability that applies to all the items in a given testlet for that student. This model discounts the testlet information less than in the partial credit model. Alternative models for testlets (Li, et al., 2006) treat the testlet effect as if it were another ability dimension in a multidimensional IRT model, with three different approaches to the general model, varying constraints on slope parameters and item discrimination parameters.

None of these models, however, directly address a critical question for computer adaptive testlets in e-learning is it enough to consider the final score achieved in a testlet or is the "path," or series of adaptive items, by which the score is achieved also important to consider? The iota model we discuss here is of importance in partially hierarchical testlets. In fully hierarchical testlest, there is only a single path through a set of items to achieve a given scoren. Partially hierarchical testlests allow more than one path through a bundle of items to achieve the same score. The iota model tests the question of how significant the pathways through adaptive testlets are. It does so with the addition of an iota pathway parameter, ι_{ijp} over pathway p, where the summation of all ι_{ijp} for the n_i paths in a given item equals zero. The difficulty of a score level achieved according to a given path becomes $\delta_{ijp,}$ where $\delta_{ijp} = \delta_i + \iota_{ijp}$. This generates the iota model shown in Equation 3 below, which will be used to model the testlet data we describe here:

$$P(X_{is} = x \mid \theta_s) = \frac{\exp \sum_{j=0}^{x} \sum_{p=0}^{n_i} (\theta_s - \delta_{ijp})}{\sum_{r=0}^{m_i} \exp \sum_{j=0}^{r} \sum_{p=0}^{n_i} (\theta_s - \delta_{ijp})} \qquad (3)$$

The likelihood function for a standard IRT model is different from an item bundle model. In a standard IRT model the likelihood function is the product of the probabilities of scores achieved on the *items* whereas for the bundle models such as the iota model it is the product of the probabilities of scores achieved on the *bundles.*

Ideally, the iota (ι_{ijp}) and ($\iota_{ijp'}$) components of the item difficulty should have relatively small differences over all p' for a given $_{ij}$, as equivalently scored item paths though the bundle should have near equal difficulties if the construct modeling assumptions hold. To model this assumption, the iota model, which is an ordered partition model (Wilson, 1992), is used. In this application, the various pathways to a single score within an item bundle are unique, therefore are given individual parameterizations even though they are scored the same. This model can be compared to the aggregate partial credit model for the data, in which the pathways are aggregated and treated as a single score, in order to gauge statistical significance of considering the iota pathway parameters individually as compared to the hierarchically more parsimonious partial credit model. The iota model is hierarchically nested within the partial credit model, and thus can be compared by a likelihood ratio chi-square test, with the difference in estimated parameters between the two models equal to degrees of freedom.

Results

Following the BEAR CAT run reported, the authors developed a clearer picture of how the "modify" option was used. First, most students did not choose to exercise the modify option very often. For most items, fewer than 20 of about 400 students (5%) who took the item choose to exercise the modify option. On a few items the number of respondents modifying the answers rose as high as 75 (19%), or as low as one student (.25%) modifying the answer.

When modifications were made, they often did not change a student's score and tended instead to simply restate the language of the answer, or in some cases extend the answer with information not requested in the original item but relevant to the situation. Rarely did modifications offer a different conception of the answer from the offered choices, but some of the modifications were valuable in identifying where questions might be improved to better reflect student thinking.

The exception to this consistency effect by modification between the selected and constructed response was when students used the modify option to in effect go back and change earlier answers in their bundle. Again, this tended to happen in two types of cases. Sometimes students said that as they moved through the bundle they had realized they couldn't be right on an earlier question in the bundle and so had changed their mind on that answer and wanted to let us know. Other students said they had misread the earlier question. In modifications in two of the bundle screen analyzed for this effect, when students said they were correcting their prior answers for either of these reasons, their modified answers were almost always an improvement on their prior answers.

CAT-based systems, including BEAR CAT, tend not to allow students to go back and change answers on prior questions are numerous. Going backwards and changing questions could alter the adaptivity flow into future questions, thus allowing students to in effect "substitute" questions they didn't want to answer by going back and changing answers such that different future items are delivered. Indeed, in most CAT systems, going back and giving the same answer would change the item delivery as well, since items are selected randomly from a pool constrained for difficulty and perhaps other aspects such as previous exposure rates in a population, though this could be handled by recalculating the student performance estimate. Perhaps the most serious problem is that astute studying of whether the questions seem to be getting more difficult or less difficult depending on a given answer can clue the student on the correct answers if going back is allowed in adaptive assessment systems.

Students were not always right when they modified answers. Modifications tended to happen in two situations. Some modifications came about from students who wanted to add more detail to the distractors they were offered, or who wanted to point out that one answer offered was completely correct but another was partially correct, as was true in many of the questions where the distractors were "ordered multiple choice" to reflect different levels of the Perspectives framework. Students who modified answers for these reasons tended to be correct in their constructed response as well as in their selected response. However, the modification option was also used frequently by students who had reached a screen that would only be offered after students had incorrectly answered in another part of the bundle. The follow-up screens were to verify that the student continued to answer incorrectly and to ascertain the knowledge with which they might be reasoning. Here students who modified answers often continued to be wrong. For instance on one bundle analyzed for this effect, the students who modified in this situation continued to be wrong seven out of eight times. Students who modified these screens, however, often offered answers that captured somewhat different incorrect reasoning than had been offered in the original distractors. These answers were especially valuable in thinking about improved item distractors.

Conclusion

When instruction is differentiated in the classroom, we often see how multiple approaches are spiraled into the curriculum. Sometimes self-selection is a differentiation strategy, in which students select among a variety of materials for learning, or options of products they can complete as assignments. In e-learning, differentiation approaches can include such self-selection, often through selecting which tutorials, representations or other materials the student would like to use, or materials they would like to generate. In this paper the authors look at a differentiation option in which students are allowed to modify assessment question answer choices. They explored the degree to which students chose to use the modification option for differentiation, and whether the assessment results changed based on their modifications.

The researchers found that students in this study exercised the option only about 5 percent of the time when it was available. Though often revealing distinctly different thinking in the proposed modification as compared to the offered answer choices, this usually did not change the student score because the new thinking often revealed incorrect understanding of the problem or task.

Instructionally, making student ideas explicit and having access to the alternate thinking gave instructors and content developers new insight into the reasoning of students. For differentiated instruction this could allow for the adaptive flow of content potentially to be improved for students. A broader range of reasoning and misunderstandings might be captured and understood by allowing students to self-suggest alternate reasoning not yet well captured in the product's assessment designs.

References

[1]. Bennett, R. E., Morley, M., Quardt, D. (2000, April). *Three Response Types for Broadening the Conception of Mathematical Problem Solving in Computerized-Adaptive Tests.* Paper presented at the National Council of Measurement Education, San Diego, CA.

[2]. Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working Inside the Black Box: Assessment for Learning in the Classroom*. London: King's College.

[3]. Black, P., & Wiliam, D. (1998). Inside the Black Box: Raising Standards through Classroom Assessment. *Phi Delta Kappan, 80*(2), 139-148.

[4]. Embretson, S., & Reise, S. P. (2000). Item Response Theory as Model-Based Measurement, *Item Response Theory for Psychologists* (pp. 40-61). Mahwah, NJ: Lawrence Erlbaum Associates.

[5]. Gifford, B. R. (2001). Transformational Instructional Materials, Settings and Economics, *The Case for the Distributed Learning Workshop*. Minneapolis, MN: The Distributed Learning Workshop.

[6]. Hall, T. (2002). *Differentiated instruction.* Retrieved November, 2006, from http://www.cast.org/publications/ncac/ncac_diffinstruc.html

[7]. Li, Y., Bolt, D. M., & Fu, J. (2006). A Comparison of Alternative Models for Testlets. *Applied Psychological Measurement, 30*(1), 3-21.

[8]. Nielsen, J. (1998). *Jakob Nielsen's Alertbox for October 4, 1998: Personalization if Over-Rated.* Retrieved Nov. 8, 2006, from http://www.useit.com /alert box/ 981 004.html.

[9]. Parshall, C. G., Spray, J., Kalohn, J., & Davey, T. (2002). Issues in Innovative Item Types, *Practical Considerations in Computer-Based Testing* (pp. 70-91). New York: Springer.

[10]. Parshall, C. G., Stewart, R., Ritter, J. (1996, April). *Innovations: Sound, Graphics, and Alternative Response Modes.* Paper presented at the National Council on Measurement in Education, New York.

[11]. Reis, S. M., Kaplan, S. N., Tomlinson, C. A., Westbert, K. L., Callahan, C. M., & Cooper, C. R. (1988). How the brain learns, A response: Equal does not mean identical. *Educational Leadership, 56*(3).

[12]. Resnick, L. B., & Resnick, D. P. (1992). Assessing the Thinking Curriculum: New Tools for Educational Reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction* (pp. 37-76). Boston, MA: Kluwer Academic Publishers.

[13]. Scalise, K. (2005). *Data Driven Content in e-Learning: Integrating Instruction and Assessment.* Paper presented at the American Association for Higher Education (AAHE) Virtual Assessment Conference, Strand 1: Promoting an Institutional Culture of Integrated Assessment, http://www.aahe.org.

[14]. Sizer, T. R. (2001). No two are quite alike: Personalized learning. *Educational Leadership, 57*(1).

[15]. Taylor, C. R. (2002). *E-Learning: The Second Wave.* Retrieved July 10, 2006,from http://www.learningcircuits. Org/2002/oct2002/taylor.html

[16]. Tomlinson, C. A. (2001). *How to differentiate instruction in mixed-ability classrooms* (2nd ed.). Alexandria, VA: ASCD.

[17]. Tomlinson, C. A., & Allan, S. D. (2000). *Leadership for differentiating schools and classrooms*. Alexandria, VA: ASCD.

[18]. Tomlinson, C. A., & McTighe, J. (2006). *Integrating Differentiated Instruction + Understanding by Design: Connecting Content and Kids*. Alexandria, VA: Association for Supervision and Curriculum Development.

[19]. Trivantis. (2005). *Present Day Custom eLearning. R*etrieved July 12, 2006, from http://www.trivantis. Com/ cust om-elearning/custom-elearning.htm

[20]. Turker, A., Görgün, I., & Conlan, O. (2006). The Challenge of Content Creation to Facilitate Personalized E-Learning Experiences. *International Journal on E-Learning, 5*(1), 11-17.

[21]. Wainer, H., Brown, L., Bradlow, E., Wang, X., Skorupski, W. P., Boulet, J., et al. (2006). An Application of Testlet Response Theory in the Scoring of a Complex Certification Exam. In D. M. Williamson, I. J. Bejar & R. J. Mislevy (Eds.), *Automated Scoring of Complex Tasks in Computer Based Testing*. Mahway, NJ: Lawrence Erlbaum Associates, Inc.

[22]. Wainer, H., Kiely, G. (1987). Item Clusters and Computerized Adaptive Testing: A Case for Testlets. *Journal of Educational Measurement, 24*, 185-202.

[23]. Willis, S., & Mann, L. (2000). *Differentiating Instruction. Curriculum Update, Education Topics, ASCD* Retrieved Nov. 2006, from http://www.ascd.org/ed_topics /cu2000win_willis.html.

[24]. Wilson, M. (1992). The Ordered Partition Model: An Extension of the Partial Credit Model. *Applied Psychological Measurement, 16*(4), 309-325.

[25]. Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Lawrence Erlbaum Assoc.

[26]. Wilson, M., & Adams, R. J. (1995). Rasch Models for Item Bundles. *Psychometrika, 60*(2), 181-198.

[27]. Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.

ABOUT THE AUTHOR

Kathleen Scalise received her Ph.D. in quantitative measurement at the University of California, Berkeley, in 2004. She is an Assistant Professor at the University of Oregon in the U.S. with work in multi-stage computer adaptive testing (CAT). She is interested in item response models for CAT, innovative item types in computer-based testing, and use of CAT algorithms for differentiated instruction in e-learning products, which is rapidly emerging in the U.S. All three approaches employ item response models to explore latent constructs and investigate effective assessment approaches. She served with the Curriculum Frameworks and Instructional Resources Division of the California Department of Education for the most recent California Science Framework for K-12 Public Schools in the U.S., and works primarily in the areas of science and mathematics assessment.