Sliding scale: A technique to optimize the assessment of knowledge level through ordering theory

Celina Byers

William A. Byers

Presented at the annual conference of the

International Personnel Management Association Assessment Council

Chicago

June, 1998

Introduction

The various learning models involved in the process of designing instruction have been the subject of much thought and writing during the past several decades. One model that has gained general acceptance, and one that has received support from eminent scholars, is hierarchical learning. People like Jean Piaget, Robert Gagne, Paul Merril, and Richard White have spent a great deal of thought, time and implementation effort on the concept, which follows the idea that parts of a domain are best learned in the proper sequence. Learning the earlier parts of the domain, in other words, enables the learner to more efficiently assimilate the later ones.

Since learners, especially adult learners, bring differing amounts and types of previous knowledge to the task of learning something new, an essential activity of hierarchical learning is to discover the knowledge level, the point in the hierarchical sequence where a learner should start. From a strictly logical perspective, each learner has a unique knowledge level because each learner has a unique genetic composition and a unique life experience. In practice, of course, groups of learners with sufficiently similar knowledge levels usually learn in groups. As computer-assisted education becomes more widespread and learners become more proficient in using technology, however, attending individual learning needs becomes more practicable. In any case, the importance of finding the learner's knowledge level, is a major consideration in the planning of instruction (Ausubel, 1968; Smith and Regan, 1993).

For more that 50 years, a succession of scholar-statisticians have pursued research that approaches hierarchical learning from an angle different from that of Piaget and Gagne. Scalogram analysis, and specifically the Guttman Scale (Guttman, 1946), a method of displaying test items by order of difficulty, served to catalyze study of the subject. A homogeneity index that measured the probability that learners will answer test items correctly and that could be used to enhance Guttman's construct was derived by Loevinger (1947); her index allowed for deviation from theoretical prediction, while Guttman's model required an unreal exactitude. Coombs (1964) and his students worked on the statistical principles behind ordering theory in the 1960s, and Krus, Bart, and Airasian (1975) continued research in the area and named the concept ordering theory. Extending the work of Loevinger (1947), Cliff (1977) made an important contribution by deriving an index he called Ct3, which indicates the reliability of a test as well as the degree of consistency its items have with the inherent learning order of the domain they represent. The progress of this line of research may show the way to systematize the design of instruction that is based on hierarchical design.

Assessment

Classical test theory holds that the single administration of a dichotomous (answers are either right or wrong) test generates for each respondent a composite, or summative, score. Each item of the test is a test in itself, or sub-test (Crocker & Algina, 1986). An important characteristic of a summative score is the lack of correspondence between score and performance level. Two learners can have matching summative scores and different performance capabilities; in other words, they can have the same score but produce incorrect answers for different items. Ordering theory, on the other hand, suggests matching the test item order to the natural learning sequence of the domain. In this case, score is an accurate measurement of performance because two respondents with the same score have responded correctly to the same items (Krus, Bart, & Airasian, 1975; Coombs, 1964; Suppes and Zinnes, 1963). Commonly, a normal bell curve is generated by the frequency distribution of a significant sample of respondents (Anastasi, 1961; Hinkle, Wiersma, & Jurs, 1994; McNemar, 1969; Thissen, 1993). This means, logically, that the sample is composed of subgroups of varying scores. When two ordering conditions obtain, namely that the test items occur in difficulty-order sequence and also match the natural learning order of the domain, ordering theory maintains that the knowledge level is defined by the first incorrect answer. Since the balance of the items are more difficult than the one that produced the wrong answer, the respondent is expected to answer them all incorrectly as well. Since the items also match the learning sequence inherent in the domain, the balance of the items represent the respondent's knowledge gap and map out the learning steps to be taken to achieve mastery of the domain. Using this idea in instructional design could permit a considerable saving by abbreviating test taking and making study more efficient.

Test and domain

An experiment by Byers (1997) illustrated how to use ordering theory principles to derive and employ the isomorphic linear relationship between the elements of a domain and the items on a test. In the first step, the order of difficulty is obtained from the proportion of correct answers. Perfect order, of course, is almost exclusively a theoretical concept; in practice, it is highly unlikely. Analysis of responses that confirm and disconfirm the theory is necessary to find the degree to which the theory is being correctly implemented. Consider, for example, a pair of test items ti5 and ti7, where ti5 has a lower difficulty index. If one of the items is answered incorrectly, the expected answer pattern is 1,0 (ti5 correct, ti7 incorrect). This pattern confirms the theory while its converse, answer pattern 0,1 (ti5 incorrect, ti7 correct) disconfirms the theory. Comparing all of the pairs of items permits the calculation of the probability that a respondent will answer as the theory predicts and makes possible appropriate adjustments to the item order (Krus, Bart, & Airasian, 1975).

In order to determine whether the ratio of confirmatory/disconfirmatory responses is significant or not, the value of an index of internal consistency is calculated. Cliff (1977) used Loevinger's (1947) index as a base and derived Ct3, an index that Cudeck (1980) subsequently showed to indicate the test's reliability as well as the degree of consistency its item order has with the learning order of the elements of the domain that are represented by the items. Cudeck's (1980) study proved that a Ct3 between .3 and .5 indicates acceptability on both counts. A test may be highly reliable and still generate a Ct3 well outside the limits of acceptability; this means that the test items do not sufficiently reflect the domain learning order. Finally, the item ordering process produces a graphical display of the test items in the order reflecting their relative difficulty and the natural learning order of the domain in question.

Two compatible research thrusts

A traditional school of thought respected by the educational community affirms that the nature of learning is hierarchical. For example, Ausubel (1963) and others (Ausubel and Robinson, 1969; Driscoll, 1994) postulate the hierarchical structure of cognition, and Gagne (1962), a long-time proponent of hierarchical learning, continues to be a strong influence (Resnick, 1973; Reiser, 1987; Merril 1987). The purpose of this paper is to meld this line of study with the work of ordering theory scholars in order to move toward the realization of making instructional design more effective.

An implementation of ordering theory

Byers (1997) demonstrated that ordering theory is feasible as a basis for instructional design. At Irma Marsh Middle School (Fort Worth, Texas), she performed an experiment in which she found a live test with an acceptable Ct3, applied ordering theory to test results, calculating interrelationships and the learning sequence of the domain. She then reordered the test items according to the learning sequence indicated by ordering theory, re-administered the same test in the new order to the same respondents, and evaluated the outcome. In the re-administration, a selected test item from each level of difficulty was presented until an wrong answer occurred, and then the rest of the items were presented in ascending order of difficulty to determine whether the right and wrong answers followed the prediction of the theory.

A correlation of 1.00 between observed and expected right answers strongly supported the idea that this method can establish knowledge level. However, a correlation of only .47 between observed and expected wrong answers indicated that stopping the test at the first wrong answer was not the best technique. A number of statistical constructs were devised to improve the accuracy of estimating knowledge level.

First, True Knowledge Level was defined as the total number of difficulty levels with at least 50% correct answers. Correlation between right and wrong observation and expectation were .97 and .98, respectively, indicating acceptable accuracy. However, economy is lost in this determination, for at least half of the items in all the difficulty levels must be administered to obtain the data to calculate the True Knowledge Level. In an attempt to maintain the capacity to abbreviate the testing process, Sliding Scale Three (SS3) and Sliding Scale Five (SS5) were elaborated. These two models were derived from the concept of moving averages, which are applied to data to smooth out the statistical irregularities (Appel and Hitschler, 1980).

To create SS3, difficulty levels 1, 2, and 3 are considered to be the first level, difficulty levels 2, 3, and 4 comprise the second, 3, 4, and 5 make up the third; this pattern repeats until all the difficulty levels are included. Of the eight possible answer patterns to each

SS3 level, (000), (001), (010), (100), (011), (101), (110), (111), the first four show a majority of wrong answers and thus indicate 0, or wrong answer, for the SS3 level. SS3 levels with answer patterns showing a majority of right answers are considered to be correct and are marked with 1. When 0 occurs as a SS3 level score, the middle of the three difficulty levels that comprise that SS3 level is considered to be the knowledge level. SS5 functions exactly the same way but uses five difficulty levels instead of three.

In statistical comparisons of observed and expected right and wrong answers, the correlations for SS3 were .98 and .88, respectively, while the same calculations for SS5 produced .97 and .93, respectively. This high correlation both of right and wrong observed versus expected indicates that ordering theory determines the learner's knowledge level with acceptable accuracy and also provides the saving of abbreviated testing.

Conclusion

It is possible and practical to employ ordering theory to find learner's knowledge levels. However, a recent study suggests that stopping the test at the first wrong answer does not ascertain knowledge level with sufficient accuracy. The application of a type of moving average, however, allows early identification of knowledge level and also provides adequate precision.

REFERENCES

Anastasi, A. (1961). Psychological testing. New York, NY: The Macmillan Company.
Ausubel, D. P. (1963). Cognitive structure and facilitation of meaningful verbal learning. Journal of Teacher Education, 14, 217-221.
Ausubel, D. P. (1968). Educational psychology: A cognitive view. New York, NY: Holt, Rinehart and Winston, Inc.
Ausubel, D. P., & Robinson, F. G. (1969). School Learning: An introduction to educational psychology. New York, NY: Holt, Rinehart and Winston, Inc.
Byers, C. (1997). Using ordering theory to establish knowledge levels. Unpublished doctoral dissertation, University of North Texas.
Cliff, N. (1977). A theory of consistency of ordering generalizable to tailored testing. Psychometrika, 42 (3), 375-399.
Coombs, C. H. (1964). A theory of data. New York, NY: John Wiley & Sons, Inc.
Crocker, L., & Algina, J. (1986). Introduction to classical & modern test theory. Orlando, FL: Harcourt Brace Jovanovich, Inc.
Cudeck, R. (1980). A comparative study of indices for internal consistency. Journal of Educational Measurement, 17 (2), 117-130.
Driscoll, M. P. (1994). Psychology of learning for instruction. Boston, MA: Allyn and Bacon.
Gagné, R. M. (1962). The acquisition of knowledge. Psychological Review, 69, 355-365.

Guttman, L. (1944). A basis for scaling qualitative data. American Sociological Review, 9, 139-150.
Guttman, L. (1946). An approach for quantifying paired comparisons and rank order. Annals of Mathematical Statistics, 17, 144-163.
Guttman, L. (1947). The Cornell technique for scale and intensity analysis. Educational Psychological Measurement, 7, 247-279.
Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1994). Applied statistics for the behavioral sciences. Boston, MA: Houghton Mifflin Company.
Krus, D. J., Bart, W. M., & Airasian, P. W. (1975). Ordering theory and methods. Los Angeles, CA: Theta Press.
Loevinger, J. A. (1947). A systematic approach to the construction and evaluation of tests of ability. Psychological Monographs, 61 (4, Whole, No. 285).
McNemar, Q. (1969). Psychological Statistics (4th ed.). New York, NY: John Wiley and Sons, Inc.
Merril, P. F. (1987). Job and task analysis. In R. M. Gagné (Ed.), Instructional technology: Foundations (pp. 141-175). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Reiser. R. A. (1987). Instructional technology: A history. In R. M. Gagné (Ed.), Instructional technology: Foundations (pp. 11-49). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Resnick, L. B. (1973). Issues in the study of learning hierarchies. Instructional Science, 2, 312-323.
Resnick, L. B. (1976). Task analysis in instructional design: Some cases from mathematics. In D. Klahr (Ed.), Cognition and Instruction (pp. 51-80). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
Smith, P. L., & Ragan, T. J. (1993). Instructional design. New York, NY: Macmillan Publishing Company.
Suppes, P, & Zinnes, J. L. (1963). Basic measurement theory. in R. D. Luce, R. R. Bush, & E. Galanter (Eds.), Handbook of mathematical psychology. New York, NY: Wiley.

Authors

Celina Byers is Assistant Professor in the Division of Technology for Training and Development (TTD) at the School of Education and the Instructional Technology Specialist at the Center for Instructional Design and Delivery (CIDD), University of South Dakota, 103 McKusick Technology Center, 414 E. Clark Street, Vermillion, SD 57069. Voice: 605.677.5247, FAX: 605.677.6518, E-mail: cbyers@usd.edu.

William A. Byers is a consultant and educational software developer, 1532 Crestview Drive, Vermillion, SD 57059. Voice: 605.624.4386, FAX: 605.624.7987, E-Mail: wbyers@usd.edu.