Psychologia, 2001, 44, 197-208

A NONPARAMETRIC ITEM ANALYSIS OF THE MOTIVATED STRATEGIES FOR LEARNING QUESTIONNAIRE — CHINESE VERSION

John SACHS, Yin kum LAW, Carol K.K. CHAN, and Nirmala RAO

The University of Hong Kong, Hong Kong S.A.R.

The quality of the Motivational Strategies for Learning Questionnaire - Chinese Version (MSLQ-CV) items was assessed with a nonparametric item analysis on a sample of 1292 Hong Kong primary and secondary students. Although a number of items within each scale needed options combined, the overall quality of these scales in discriminating between trait levels was good. Multivariate analysis on the maximum likelihood estimates of the scale scores showed significant grade and gender differences for some scales, but the effect sizes were small. Correlational analysis of these scales supported earlier research results, and a regression analysis found that four scales significantly predicted participants' scores on a standardized measure of Chinese language achievement, although the amount of variance explained was less than 5 percent. The advantages of nonparametric item analysis over traditional sample dependent omnibus measures of item and scale quality are discussed.

Key words: MSLQ, MSLQ-CV, nonparametric item analysis, TestGraf, item response theory

The Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich & De Groot, 1990) is a self-report inventory, which measures three motivational (Self-Efficacy, Intrinsic Value, and Test Anxiety) and two self-regulated learning (Strategy Use and Self-Regulation) components. Although the psychometric properties of the MSLQ have been reported in Western cultures (McClendon, 1996; Pintrich & De Groot, 1990; Pintrich, Smith, Garcia, & McKeachie, 1993) and in Asian cultures for the Chinese version (MSLQ-CV; Rao & Sachs, 1999; Rao, Moely, & Sachs, 2000), typically this has involved reporting sample dependent omnibus measures of scale quality such as the average item difficulty (or endorsement), item-total correlations, reliability, and factor loadings all of which ignore how scale item performance varies as a function of the trait being measured. Furthermore, the univariate analyses reported in these studies have relied on the MSLQ scale scores which do not take into account item and item option functioning and, therefore, are not optimal measures of the underlying trait.

However, more powerful item-based methods of measurement that are not dependent on the individuals sampled or on the particular items administered and which can model the nonlinear relationship between the probability of endorsing an item and the underlying latent trait level are available. These are the item response

All correspondence should be sent to John Sachs, Department of Education, The University of Hong Kong, Pokfulam Road, Hong Kong SAR (E-mail: jsachs@hkucc.hku.hk).

theory models (IRT) or *latent trait theory* models (Hambleton & Cook, 1977; Hambleton, 1979; Lord & Novick, 1968; Lord, 1980). Typically, most applications of IRT models assume a unidimensional latent trait space in which only a single latent trait is needed to account for an individual's item response.

The most commonly used IRT models are the 1-, 2-, and 3-parameter models for dichotomously scored achievement test items (Hambleton & Swaminathan, 1985; Lord, 1980). These parametric IRT models model the relationship between an examinee's ability level and his/her probability of answering an item correctly with a logistic function. This nonlinear regression function between item score and trait or ability level is referred to variously as an item characteristic curve (ICC), item response curve (IRC), item response function (IRF), or response characteristic curve (RCC).

Extensions of the IRT models for dichotomously scored items to polychotomous (or polytomous) items with ordered categories such as Likert scales have been made by Samejima (1969) – graded response model, Masters (1982) – partial credit model, and Andrich (1978a, b, c) – rating scale model. Some applications of these dichotomous and polychotomous IRT models to psychological and attitudinal inventories include (a) attitude scale translations (Hulin, Dragrow, & Komocar, 1982); (b) evaluation of a job descriptive index (Parsons & Hulin, 1982); (c) development of a loneliness scale (de Jong-Gierveld & Kamphuis, 1985); (d) evaluation of items measuring social norms, conflict, political view, and life satisfaction (Thissen & Steinberg, 1988); (e) analysis of a children's depression inventory (DeRoos & Allen-Meares, 1992), (f) the analysis of an approaches to study inventory (Waugh & Addison, 1998), and (g) a comparison of the structural properties of two versions of a psychopathology instrument (Cooke, Michie, & Hart, 1999).

Although these studies have shown convincingly the power of IRT for the analysis of psychological and attitudinal items, the application of parametric IRT models by the average practitioner is still difficult. First, a fair amount of statistical and mathematical sophistication is required to understand these models. Second, the programs to fit these models such as Bilog (Mislevy & Bock, 1982) for dichotomous items and Multilog (Thissen, 1988) for polychotomous items can be difficult to use and interpret. Third, large sample sizes, often 500 to 1000 subjects (Lord, 1980; Reise & Yu, 1990; Swaminathan & Gifford, 1983), are needed for accurate parameter estimates. And, fourth, unlike achievement test items, personality and attitudinal inventory items have typically not been developed with a parametric IRT model in mind, thus the fit of a logistic model to such items may not be optimal (Ramsay, 1997). However, nonparametric kernel-smoothing techniques that can model asymmetry in the item-trait regression, that are easy to use and interpret, and that do not require large samples, have been developed by Ramsay (1991, 1993, 1997).

Nonparametric IRT

Although nonparametric IRT methods typically assume, like their parametric counterparts, that only one latent trait is being measured (unidimensionality), they make no restrictive assumptions about the shape of the curve(s) linking the probability of endorsing an item or item options to the underlying trait and present this curve(s) in graphical form. Therefore, computational problems associated with parameter estimation are avoided as are the need for statistical and mathematical sophistication on the part of the user in interpreting item and option parameters associated with parametric IRT models (Ramsay, 1997). Furthermore, the nonparametric approach can account for response characteristic curve asymmetry, as noted earlier, and also for nonmonoticity (Ramsay, 1991), thus achieving a better degree of fit to the data. And like parametric IRT approaches, detailed information regarding item, option, and scale effectiveness is supplied along with maximum likelihood estimates of the trait level, which include unique standard error estimates.

Since the characteristics of MSLQ items have been studied only with traditional sample dependent measures, the purpose of this study was to assess the quality of the items in each scale with nonparametric IRT methods and to obtain maximum likelihood estimates of participants' trait levels. Specifically we were interested in evaluating (a) the relative effectiveness of the items and item options within each scale at different levels of the trait being measured (item and option effectiveness), (b) the ability of MSLQ scale items to detect individual differences on the underlying trait (scale discrimination), (c) the effect of gender and grade level on participants MSLQ scale scores, and (d) the relationship of these scales to academic achievement.

Method

Participants:

Participants were 1,292 Hong Kong Chinese primary and secondary school students – 530 Primary 5 (P5: $Mdn_{age} = 11$), 533 Primary 6 (P6: $Mdn_{age} = 12$), and 229 Secondary 1 (S1: $Mdn_{age} = 13$). Of these students, 530 were boys and 695 were girls, with 12 missing gender values.

Measures:

Motivated Strategies for Learning Questionnaire - Chinese Version (MSLQ-CV). The Motivational Strategies for Learning Questionnaire (MSLQ; Pintrich & De Groot, 1990) is a 44 item self-report inventory that uses a 7-point Likert scale (1 = *not at all true of me* to 7 = *very true of me*) to measure students' motivational orientation and self-regulated learning strategy use. This version of the MSLQ taps three motivational (Self-Efficacy, Intrinsic Values, and Test Anxiety) and two self-regulated learning strategy (Cognitive Strategy Use and Self-Regulation) components or scales. Some slight rewording of the MSLQ-CV was made to make it suitable for upper primary students.

Further adjustment involved the four self-regulated learning strategy items – items 26, 27, 37, and 38 – the only MSLQ items that need to be reflected (i.e., reverse coded). Rao and Sachs (1999) found that these items formed a separate "Methods" scale, which reflected difficulties Chinese respondents had in dealing with reverse-coded items. We also found a separate "Methods" scale and excluded these items from further analysis. A complete list of the MSLQ-CV items can be found in the Appendix.

The Hong Kong Achievement Tests (Chinese) – HKAT-C. This standardized Chinese language achievement test was developed by The Hong Kong Education Department to assess the academic performance of Hong Kong's primary and lower secondary school pupils. It was included to examine the extent to which the MSLQ-CV predicted student performance.

Analytic Method:

Although the unidimensionality assumption associated with most IRT methods seems overly restrictive for personality and attitudinal scale items, Hambleton, Swaminathan, and Rogers (1991) have noted that all that is really required for the unidimensional assumption to be adequately met by a set of items is the presence of a dominant eigenvalue. And Reckase (1979) has suggested that this dominant eigenvalue need account for just over 20 percent of the variance in the item pool for the underlying assumption of unidimensionality to be reasonably met.

Using the above criteria, the principal component analysis of each scale showed that the dominant eigenvalues for the Self-Efficacy (9 items, alpha = 0.88), Intrinsic Values (9 items, alpha= 0.81), Test Anxiety (4 items, alpha = 0.77), Cognitive Strategy Use (12 items, alpha = 0.84), and Self-Regulation (6 items, alpha = 0.74) scales accounted for 44.70%, 34.90%, 46.14%, 31.20%, and 33.25% of the variance, respectively. Thus the five scales were regarded as essentially unidimensional.

The nonparametric item analysis was conducted with TestGraf (Ramsay, 1993), which uses nonparametric kernel-smoothing techniques (Altman, 1992; Ramsay, 1991, 1997) to fit item and option response characteristic curves. It is generally recommended (Ramsay, 1993) that TestGraf be used with 20 or more items and a sample size of about 100 or more respondents. The actual item requirements, however, are that the number of items "or" choices exceed 20. So four 5–option items would give 20 choices and technically meet the above requirement. And since IRT models have been fit to data involving only three polychotomous items (Thissen & Steinberg, 1988) using Multilog (Thissen, 1988), Ramsay has suggested that as few as three items could be used with TestGraf too (personal communication, 16 November 2000) provided that the sample size is large. Since our sample size was quite large and since the number of choices for the shortest scale – Test Anxiety – equaled 28 (4 items by 7 options), the use of TestGraf with our data presented no special problems.

Results

To ensure that the most accurate maximum likelihood estimates of the trait values were computed, we specified 101 display values (the maximum allowable with TestGraf), which controls the number of equally spaced values used to display the latent trait. Also, we set the kernel smoothing parameter, which controls how smooth the item and option characteristic curves appear, at 0.35 since this value was shown to work well with a psychological scale analyzed by Ramsay (1993).

Item and Option Effectiveness

The degree of effectiveness of an item at any point on the trait is indicated by the rate of change of the response characteristic curve (Santor, Ramsay, & Zuroff, 1994). Fig. 1 shows the response characteristic curve (RCC) for item 13 (I think I will receive a good grade in this class) of the Self-Efficacy scale. The vertical error bars shown on the curve indicate the 95% *point-wise confidence limits* for the true curve value (item response), and the longer vertical dashed lines (5%, 25%, 50%, 75%, and 95% lines shown) indicate the percentage of respondents that fell below a given trait level. The RCC in Fig. 1 increases rapidly with increasing trait level over the whole trait range and since the error bars are small it indicates that the curve has been precisely estimated for these trait levels. Thus item 13 was effective in discriminating between trait levels.

However, item effectiveness is ultimately dependent on option effectiveness. For items using ordered-response categories, such as here, the probability of endorsing each option category should be dominant over a limited range of the trait, reflecting the rank ordering of the response categories. The item option characteristic curves

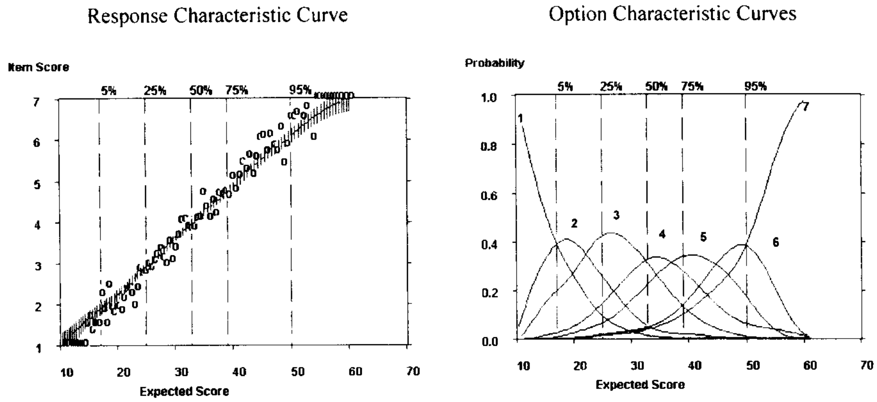

Fig. 1. Response characteristic curve and option characteristic curves for Self-Efficacy Item 13.

(OCC) can reveal whether or not the ideal ordered pattern is obtained. If some OCCs are virtually indistinguishable or if their endorsement probability is never dominant, then differential weighting of these options is not appropriate, and they probably should be combined if accurate estimates of the trait are desired (Santor & Ramsay, 1998). The OCCs for item 13 (see Fig. 1) reflect a nearly ideal ordering of the response options thus justifying the prior option weights.

Unfortunately, space limitations do not permit displaying the OCCs for the remaining 43 items. But inspection of their OCCs revealed that slightly more than half the items within each of the five MSLQ scales needed to have one or more of their options combined since some OCCs were virtually identical or were completely contained within other OCCs. That is, item options that needed to be combined were ones whose OCC never dominated the response probability over an appropriate range of the trait. The OCCs for items that had options combined were reexamined to ensure that the new OCCs reflected the ideal rank ordering of the new item option weighting. Although the maximum possible total score for each scale was reduced because of item option combining, this process ensured that the most accurate estimates of the trait were obtained for each scale.

Scale Quality

Traditional estimates of scale quality such as internal-consistency reliability and the *standard error of measurement* (SEM) are flawed because they are omnibus sample specific measures that do not vary as a function of the trait or ability level (Hambleton & Swaminathan, 1985; Ramsay, 1993).

IRT, however, offers a direct measure of test (scale) quality – the information function. The test information function is such a superior measure of scale quality because it uses all the information in the test or scale items and because it varies as a function of underlying trait or ability level (Santor & Ramsay, 1998).

TestGraf plots the average information function so that scales with different numbers of items can be compared. It also plots the standard error function, which is just the square root of the reciprocal of the information function. For presentation purposes, the latter is probably preferable since the units should be more familiar to practitioners. Therefore, the standard error functions for the five MSLQ scales are presented in Fig. 2.

Inspection of Fig. 2 shows that all five standard error functions had two troughs at either end of the trait range and a peak near the centre of the trait range. Thus the scales provided slightly more accurate trait measures for some of the lower and higher expected scores and slightly less accurate trait measures for those with expected scores near the centre. Consequently, the MSLQ scales do not measure their respective traits evenly over the whole range of trait values.

The maximum range of the standard error functions was about 1 and the minimum was about 0.4. This means that the item characteristic curves for these items had slopes that were in the range of about .25 to 1.10 indicating that on the whole the MSLQ items were fairly discriminating. The standard error function for the Self-Efficacy scale was the most variable and showed the largest standard errors, so its items were the least discriminating and informative, while the Test Anxiety scale had the smallest standard errors with a range of about 0.4. Ranking these scales in terms of their standard error functions from the lowest to highest overall standard errors gave us the following ordering: (1) Test Anxiety, (2) Self-Regulation, (3) Intrinsic Value, (4) Cognitive Strategy Use, and (5) Self-Efficacy. It is interesting to note that had we used the internal consistency reliability estimates presented earlier as a measure of scale quality the rank ordering of these scales would have been almost exactly opposite with Self-Efficacy rated first.

Since there was considerable overlap in the ranges of these standard error functions, with the exception of perhaps Self-Efficacy and Cognitive Strategy Use, and because the standard errors for the trait estimates were fairly small, with the exceptions just noted, we felt confident in using these measures to examine gender and grade differences.

Univariate and Multivariate Scale Analyses

Because the maximum likelihood estimates of trait level computed by TestGraf – the expected scores – are based on a weighted sum of the individual option characteristic curves, they make use of all the information contained in an item and provide much more accurate estimates of the trait (Santor, Ramsay, & Zuroff, 1994) than conventional sum scores. Consequently, we used the expected scores in the subsequent scale level analyses.

Gender and Grade differences. A two-way MANOVA was run with gender and grade level (P5, P6, S1) as independent variables and the five MSLQ scales as dependent variables. Statistically significant multivariate F values ($p < .001$) were observed for gender, grade and grade by gender. But since the two-way interaction was statistically significant we restricted our attention to the univariate F's associated with this interaction to determine which scales were involved.

The following F values were observed for each scale: (a) Self-Efficacy, $F(2, 1274)$

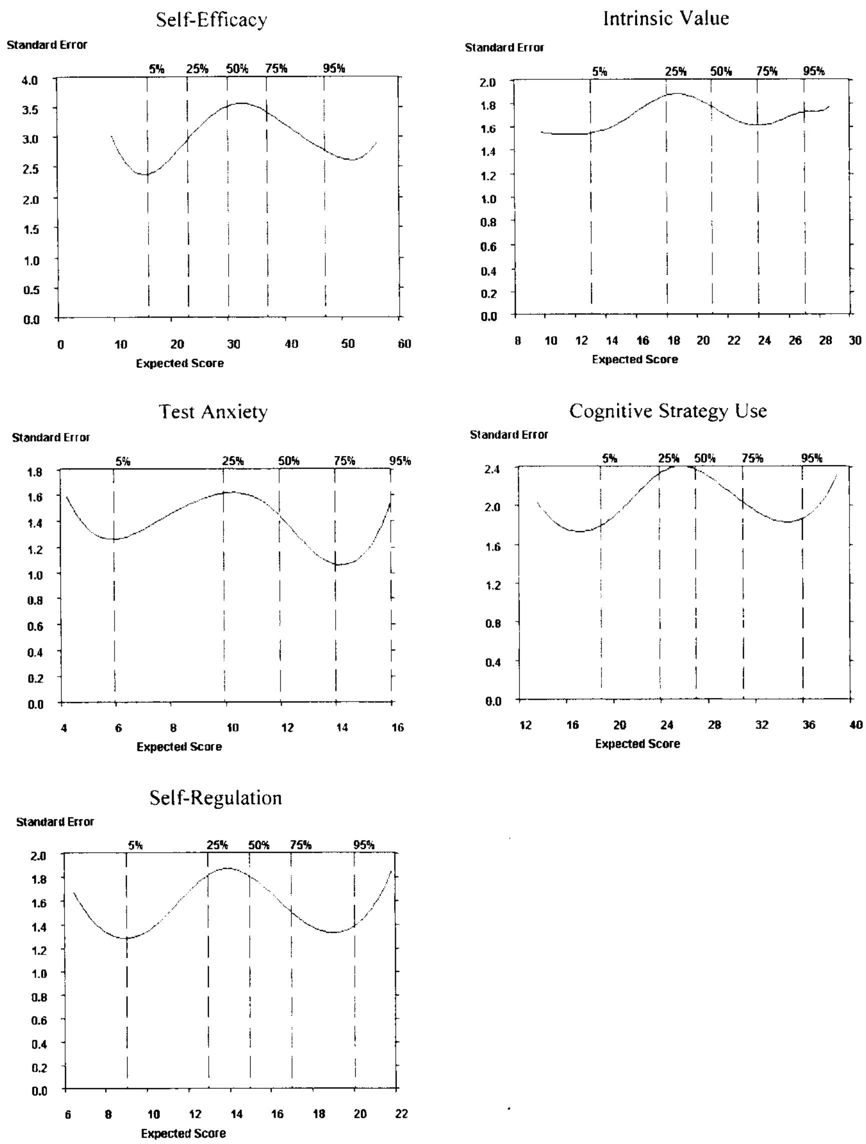

Fig. 2. Standard error functions for MSLQ-CV Scales.

$= 6.07, p < .01$; (b) Intrinsic Value, $F(2, 1274) = 4.67, p < .01$; (c) Test Anxiety, $F(2, 1274) = 2.62$, $p > .05$; (d) Cognitive Strategy Use, $F(2, 1274) = 3.29, p < .05$; and (e) Self-Regulation, $F(2, 1274) = 1.93$, $p > .05$. Although Self-Efficacy and Intrinsic Value were statistically significant at the .01 level (our test value), the strength of association as measured by eta-squared (η^2) was only .009 for Self-Efficacy and .007 for Intrinsic Value. So less than 1% of the variance in these scales was accounted by gender and grade level. Even by conventional standards (see Cohen,

Table 1. Summary Statistics and Zero-Order Correlationds Between MSLQ-CV Scales and HKAT-C

	1	2	3	4	5	6	*Mean*	*SD*
1 HKAT-C	1.000						71.1	14.52
2 Self-Efficacy	.089**	1.000					3.7	1.11
3 Intrinsic Value	.088**	.524***	1.000				4.8	1.00
4 Test Anxiety	.147***	.203***	−.043	1.000			4.4	1.43
5 Cognitive Strategy	.006	.379***	.657***	−.057	1.000		4.5	1.02
6 Self-Regulation	.104**	.346***	.592***	.000	.678***	1.000	4.5	1.26

Note: $N = 1292$. $^{*}p < .05$, $^{**}p < .01$, $^{**}p < .001$.

1977) such small effect sizes would not be considered substantively meaningful. Therefore, there seemed little point in examining such small mean differences in any detail.

Zero-order correlations between scales and Chinese language achievement. The inter-correlations between the MSLQ scales and their correlations with HKAT-C are shown in Table 1. Inspection of Table 1 shows that Self-Efficacy correlated significantly ($p < .001$) with all the other MSLQ scales, while Intrinsic Value, Cognitive Strategy Use, and Self-Regulation correlated significantly ($p < .001$) with all the other MSLQ scales except for Test Anxiety. The largest correlations were between Cognitive Strategy Use and Self-Regulation ($r = .678$, $p < .001$), Cognitive Strategy Use and Intrinsic Value ($r = .657$, $p < .001$), Intrinsic Value and Self-Regulation ($r = .592$, $p < .001$), and Intrinsic Value and Self-Efficacy ($r = .524$, $p < .001$). All MSLQ scales except for Cognitive Strategy Use correlated significantly ($p < .01$) with HKAT-C. However, Test Anxiety ($r = .147$, $p < .001$) and Self-Regulation ($r = .104$, $p < .001$) showed the highest correlations with HKAT-C.

Regression of HKAT-C on MSLQ scales. A stepwise regression of HKAT-C on the MSLQ scales resulted in a statistically significant regression equation, $F(4, 1287) = 3.29$ ($p < .001$), that retained four predictors – Test Anxiety ($\beta = .14$; $t = 4.93$, $p < .001$), Self Regulation ($\beta = .15$; $t = 3.91$, $p < .001$), Cognitive Strategy Use ($\beta = -.15$; $t = 3.54$, $p < .001$), and Intrinsic Value ($\beta = .09$; $t = 2.36$, $p < .05$). The negative beta weight for the Cognitive Strategy Use scale is a classic example of a suppression (see Lord & Novick, 1968) since Cognitive Strategy Use had a near zero correlation with the criterion (HKAT-C) and large correlations with two predictors – Intrinsic Value ($r = .657$) and Self Regulation ($r = .678$). Although the regression equation was significant, these four MSLQ scales accounted for only 4.2% of the variance in HKAT-C scores.

Discussion

The purpose of this study was to assess the quality of the MSLQ-CV scales using a non-parametric IRT method and to examine the effect of gender and grade level on these scales and also to determine the relationship of these scales to educational achievement as measured by the HKAT-C.

Scale quality. Although the options for a number of items within each of the MSLQ-CV scales needed to be combined, the quality of the scales was fairly good overall. This was reflected in the item response characteristic curves that increased rapidly with increasing trait levels so that items discriminated well and in the item option characteristic curves, which were sensitive to changes in the trait level thus reflecting the rank ordering of the item options. Furthermore, the standard error functions showed that these scales provided relatively good (small errors) estimates of the trait over the whole trait range.

Gender and grade differences. The effect of gender and grade level on the MSLQ-CV scale scores and the relationship of these scales to the HKAT-C were examined using the maximum likelihood estimates of the trait level provided by TestGraf. As noted earlier, these expected scores are superior to traditional sum scores since they make use all the information contained in the item responses.

Although we did observe a statistically significant gender by grade interaction, it did not represent a substantively meaningful association. However, our results, at least for gender, are somewhat consistent with those reported by Pintrich and De Groot (1990) since they found no gender differences for the two Cognitive scales or the Intrinsic Value scale. And the statistically significant differences they did report for the Self-efficacy and Test Anxiety scales had small effect sizes (just around .50). Because neither Pintrich and De Groot nor Rao et al. (2000) looked at grade differences, we cannot make a valid comparison between their work and ours. But Rao et al. did report two-year longitudinal data results, and although they did observe mean differences in the MSLQ-CV scale scores over these two years, these differences did not suggest any pattern of increasing or decreasing scale scores from one year to the next. And even though our analysis only involved cross-sectional data, our findings appear similar to those of Rao et al. across grades levels.

The probable reason we did not observe meaningful grade differences is that the median age difference from one grade to the next was only one year. Thus the age levels we were dealing with were just too homogeneous for any substantively meaningful age related difference to manifest themselves.

Zero-order correlations between scales and Chinese language achievement. Both the pattern and the magnitude of the zero-order correlations we observed between the Motivation and Cognitive scales were basically similar to those reported by Pintrich and De Groot (1990). We also found that Test Anxiety and Self-Regulation showed weak but significant correlations with academic achievement as measured by the HKAT-C although our correlations were smaller than those reported by Pintrich and De Groot. But our results are somewhat at variance with those reported in Rao et al. (2000) since they observed no relationship between academic achievement and either the Motivation scales or Cognitive scales. This may be because Rao et al. used the Hong Kong Certificate of Education (HKCEE) mathematics as their measure of academic achievement, while we used the HKAT-C, a measure of language achievement. Thus in our study four of the MSLQ-CV scales were able to account for a small amount of the variance (4.2%) in the HKAT-C scores.

The explanation for such a poor performance of the MSLQ-CV scales in predicting HKAT-C scores is similar to that given by Rao et al. (2000) for HKCEE-mathematics, which is that Hong Kong children being a product of an exam oriented system all know and are motivated to use the same cognitive strategies that are not necessarily linked to the standard tests of achievement they are exposed to. Therefore, whenever such uniformity in study strategies exists among respondents, one would expect the MSLQ to be a poor predictor of academic achievement.

Conclusion

IRT methods for attitude scale analysis especially the nonparametric kernel smoothing techniques have obvious benefits over traditional classical test theory methods. First, the nonlinear relationship between item response and trait level can be modeled so that item and item option quality can be ascertained by examining the item and option characteristic curves that vary as a function of the trait level. Second, traditional univariate and multivariate scale analyses can be improved by using the more accurate maximum likelihood scale score estimates. And third, an unambiguous measure of test or scale quality that varies as a function of the trait level and uses all the information in the items is available.

Reference Note

Ramsay, J. O. 1993. *TestsGraf: A program for the graphical analysis of multiple choice test and questionnaire data.* Unpublished manuscript, McGill University, Montréal, Québec, Canada.

REFERENCES

Altman, N. S. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician,* **46**, 175–185.

Andrich, D. 1978a. A rating formulation for ordered response categories. *Psychometrika,* **43**, 561–573.

Andrich, D. 1978b. Applications of a psychometric rating model to ordered categories which are scored with successive integers. *Applied psychological measurement,* **2**, 581–594.

Andrich, D. 1978c. Scaling attitude items constructed and scored in the Likert tradition. *Educational and Psychological Measurement,* **38**, 665–680.

Cohen, J. 1977. *Statistical power analysis for the behavioral sciences* (Rev. ed.). New York: Academic Press.

Cooke, D. J., Michie, C., & Hart, S. D. 1999. Evaluating the screening version of the Hare Psychopathy Checklist - Revised (PCL:SV): An item response theory analysis. *Psychological Assessment,* **11**, 3–13.

de Jong-Gierveld, J., & Kamphuis, F. 1985. The development of a Rasch-type loneliness scale. *Applied Psychological Measurement,* **9**, 289–299.

DeRoos, S. Y., & Allen-Meares, P. 1992. Rasch analysis: Its description and use analyzing the Children's Depression Inventory. *Journal of Social Service Research,* **16**, 1–17.

Hambleton, R. K. 1979. Latent trait models and their applications. In R. Traub (Ed.), *Methodological developments: New directions for testing and measurement (No. 4).* San Fransisco: Jossey-Bass.

Hambleton, R. K., & Cook, L. L. 1977. Latent trait models and their use in the analysis of educational test data. *Journal of Educational Measurement,* **14**, 75–96.

Hambleton, R. K., & Swaminathan, H. 1985. *Item response theory: Principles and applications.* Boston: Kluwer-Nijhoff.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. 1991. *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Hulin, C. L., Drasgow, F., & Komocar, J. 1982. Applications of item response theory to analysis of attitude scale translations. *Journal of Applied Psychology,* **67**, 818–825.

Lord, F. M. 1980. *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Lord, F. M., & Novick, M. R. 1968. *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

McClendon, R. C. 1996. Motivation and cognition of preservice teachers: MSLQ. *Journal of Instructional Psychology,* **23**, 216–220.

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

Mislevy, R. J., & Bock, R. D. 1982. *BILOG 3: Item analysis and test scoring with binary logistic models.* Moorsville, IN: Scientific Software, Inc.

Parsons, C. K., & Hulin, C. L. 1982. An empirical comparison of item response theory and hierarchical factor analysis in applications to the measurement of job satisfaction. *Journal of Applied Psychology,* **67**, 826–834.

Pintrich, P. R., & De Groot, E. V. 1990. Motivational and self-regulated components of classroom academic performance. *Journal of Educational Psychology,* **82**, 33–40.

Pintrich, P. R., Smith, D. A., Garcia, T., & Mckeachie, W. J. 1993. Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement,* **53**, 801–813.

Ramsay, J. O. 1991. Kernel-smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika,* **56**, 611–630.

Ramsay, J. O. 1997. A functional approach to modeling test data. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 381–393). New York: Springer.

Rao, N., & Sachs, J. 1999. Confirmatory factor analysis of the Chinese version of the Motivated Strategies for Learning Questionnaire. *Educational and Psychological Measurement,* **59**, 1016–1029.

Rao, N., Moely, B. E., & Sachs, J. 2000. Motivational beliefs, study strategies, and mathematics attainment in high- and low-achieving Chinese secondary school students. *Contemporary Educational Psychology,* **25**, 287–316.

Reckase, M. D. 1979. Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics,* **4**, 207–230.

Reise, S. P., & Yu, J. 1990. Parameter recovery in the graded response model using Multilog. *Journal of Educational Measurement,* **27**, 133–144.

Samejima, F. 1969. Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph,* No. 17.

Santor, D. A., & Ramsay, J. O. 1998. Progress in the technology of measurement: Applications of item response models. *Psychological Assessment,* **10**, 345–359.

Santor, D. A., Ramsay, J. O., & Zuroff, D. C. 1994. Nonparametric item analyses of the Beck Depression Inventory: Evaluating gender item bias and response option weights. *Psychological Assessment,* **6**, 255–270.

Swaminathan, H., & Gifford, J. A. 1983. Estimation of parameters in the three-parameter latent trait model. In D. Weiss (Ed.), *New horizons in testing* (pp. 13–30). New York: Academic Press.

Thissen, D. 1988. *Multilog version 5.1: Multiple, categorical item analysis and test scoring using item response theory.* Mooresville, IN: Scientific Software, Inc.

Thissen, D., & Steinberg, L. 1988. Data analysis using item response theory. *Psychological Bulletin,* **104**, 385–395.

Waugh, R. F., & Addison, P. A. 1998. A Rasch measurement model analysis of the Revised Approaches to Study Inventory. *British Journal of Educational Psychology,* **68**, 95–112.

(*Manuscript received March 5, 2001; Revision accepted August 10, 2001*)

Appendix

Items and scales of the Motivated Strategies for Learning Questionnaire - Chinese Version

Self-Efficacy

2 . Compared to other students in this class I expect to do well.
6 . I am certain that I can understand the ideas taught in my classes.
8 . I expect to do very well in school.
9 . Compared with others in this class, I think I am a good student.
11. I am sure I can do an excellent job on the class assignments and homework.
13. I think I will receive good grades in my exams.
16. My study skills are excellent compared with others in this class.
18. Compared with other students in this class I think I know a great deal about the subjects I am studying.
19. I know that I will be able to learn the materials for the tests and exams.

Intrinsic Value

1 . I prefer class work that is challenging so I can learn new things.
4 . It is important for me to learn what is being taught in school.
5 . I like what I am learning in school.
7 . I think I will be able to use what I learn in one subject in another.
10. I often do more than is required of me for homework assignments.
14. Even when I do poorly on a test or exam I try to learn from my mistakes.
15. I think that what I am learning in school is useful for me to know.
17. I think that what we are learning in school is interesting.
21. Understanding the subject is important to me.

Test Anxiety

3 . I am so nervous during a test that I cannot remember facts that I have learned
12. I have an uneasy, upset feeling when I take a test or exam.
20. I worry a great deal about tests and exams.
22. When I take a test I think about how poorly I am doing.

Cognitive Strategy Use

23. When I study for a test, I try to put together the information from class and from the textbook.
24. When I do homework, I try to remember what the teacher said in class so I can answer the question correctly.
28. When I study I put important ideas into my own words.
29. I always try to understand what the teacher is saying even if it doesn't make sense.
30. When I study for a test I try to remember as many facts as I can.
31. When studying, I copy my notes over to help me remember materials.
34. When I study for a test I practice saying the important facts over and over to myself.
35. Before I begin studying I think about the things I will need to do to learn.
39. When I am studying a topic, I try to make everything fit together.
41. When I read material for my classes, I say the words over and over to myself to help me remember.
42. I outline the chapters in my book to help me study.
44. When I am studying I try to connect the things I am reading about with what I already know.

Self-Regulation

25. I ask myself questions to make sure I know the material I have been studying.
32. I work on practice exercises and answer end of chapter questions even when I don't have to.
33. Even when study materials are dull and uninteresting, I keep working until I finish.
35. I use what I have learned from old homework assignments and the textbook to do new assignments.
40. When I am studying I stop once in a while and go over what I have read.
43. I work hard to get a good grade even when I do not like a class.