British Journal of Educational Psychology (2009), 79, 547–576

The British Psychological Society

www.bpsjournals.co.uk

Modern psychometrics for assessing achievement goal orientation: A Rasch analysis

Krista R. Muis[1]*, Philip H. Winne[2] and Ordene V. Edwards[3]

[1]Department of Educational and Counselling Psychology, McGill University, Montreal, Quebec, Canada

[2]Faculty of Education, Simon Fraser University, Burnaby, Canada

[3]Department of Educational Psychology, University of Nevada–Las Vegas, Las Vegas, USA

Background. A program of research is needed that assesses the psychometric properties of instruments designed to quantify students' achievement goal orientations to clarify inconsistencies across previous studies and to provide a stronger basis for future research.

Aim. We conducted traditional psychometric and modern Rasch-model analyses of the Achievement Goals Questionnaire (AGQ, Elliot & McGregor, 2001) and the Patterns of Adaptive Learning Scale (PALS, Midgley *et al.*, 2000) to provide an in-depth analysis of the two most popular instruments in educational psychology.

Samples and methods. For Study 1, 217 undergraduate students enrolled in educational psychology courses participated. Thirty-four were male and 181 were female (two did not respond). Participants completed the AGQ in the context of their educational psychology class. For Study 2, 126 undergraduate students enrolled in educational psychology courses participated. Thirty were male and 95 were female (one did not respond). Participants completed the PALS in the context of their educational psychology class.

Results. Traditional psychometric assessments of the AGQ and PALS replicated previous studies. For both, reliability estimates ranged from good to very good for raw subscale scores and fit for the models of goal orientations were good. Based on traditional psychometrics, the AGQ and PALS are valid and reliable indicators of achievement goals. Rasch analyses revealed that estimates of reliability for items were very good but respondent ability estimates varied from poor to good for both the AGQ and PALS. These findings indicate that items validly and reliably reflect a group's aggregate goal orientation, but using either instrument to characterize an individual's goal orientation is hazardous.

**Correspondence should be addressed to Dr Krista R. Muis, McGill University, 3700 McTavish Street, Montreal, Quebec, H3A 1Y2, Canada (e-mail: Krista.Muis@McGill.ca).*

DOI:10.1348/000709908X383472

Theoretical framework

Achievement goal theory is a prominent area of theoretical and empirical inquiry in educational psychology. Achievement goal orientations are a 'set of behavioral intentions that determine how students approach and engage in learning activities' (Meece, Blumenfeld, & Hoyle, 1988, p. 514). Goal orientations that students establish for themselves are claimed to be consistent across a range of academic situations and learning tasks (e.g. Ames, 1992) although there is some evidence to the contrary (Edwards & Muis, 2008; Fryer & Elliot, 2007; Winne, Muis, & Jamieson-Noel, 2006). Initially, several motivational theorists proposed an achievement goal orientation framework that included two types of goals considered particularly relevant in an achievement setting - a mastery goal orientation and a performance goal orientation (e.g. Dweck & Leggett, 1988; Meece, 1991).

Each goal orientation was theorized to provide a unique perceptual-cognitive framework in achievement settings, and early research demonstrated that each goal orientation led to a distinct pattern of processes and outcomes (Ames, 1992; Ames & Archer, 1988; Dweck, 1999). For example, a high level of mastery goal orientation was theorized to correlate with a focus on learning and on developing competence (an intra-personal perspective). In contrast, a performance orientation was predicted to covary with a need to demonstrate competence or avoid demonstrations of incompetence (an interpersonal perspective where performance is compared to others). Although positive relationships have generally been found between a mastery goal orientation, performance outcomes (e.g. Butler, 1993), self-efficacy (Phillips & Gully, 1997) and knowledge (e.g. Fisher & Ford, 1998), results linking performance orientation and various outcomes have been inconsistent. For example, Ford, Smith, Weissbein, Gully, and Salas (1998) found that a performance goal orientation had a negative relationship to task performance. VandeWalle, Brown, Cron, and Slocum (1999) found no statistically detectable relationship between a performance goal orientation and task performance. Hoover, Steele-Johnson, Beauregard, and Schmidt (1999) reported a positive relationship between these two constructs.

Theorists suggested that such inconsistencies resulted from a dichotomous model of goal orientation (e.g. Elliot, 1999; Elliot & Church, 1997; VandeWalle, 1997). Specifically, performance goal orientation encompassed a desire to obtain favourable judgments regarding one's ability plus a desire to avoid unfavourable ones. Accordingly, Middleton and Midgley (1997), Pintrich (2000a, 2000b, 2000c), and Elliot and colleagues (e.g. Elliot & Church, 1997; Elliot & Harackiewicz, 1996) expanded the performance construct to distinguish approach and avoidance motivation. This yielded two performance goal orientations: performance approach, whereby a learner strives to demonstrate aptitude and seek favourable judgments; and performance avoid rooted in a fear of failure (Elliot & Church, 1997). As Midgley *et al.* (2000) observed, recent research supports this new three-goal framework (a mastery goal, and two performance goals; Elliot & Church, 1997; Middleton & Midgley, 1997; VandeWalle, 1997).

More recently, motivational theorists (e.g. Elliot & McGregor, 2001; Pintrich, 1999, 2000a, 2000b) posited four fundamental goal orientations in which each of mastery and performance goal orientation should differentiate an approach from an avoidance direction. From a mastery-avoid orientation, a learner's goal is to avoid failure from an intra-personal perspective (relative to oneself) versus in comparison to others. Specifically, self-evaluated incompetence is the focus. A mastery-avoid oriented learner, for example, may strive to avoid misunderstanding or failing to learn course material, strive not to forget what has been learned, or try not to lose one's physical or intellectual capabilities (Elliot & McGregor, 2001). Conceptually, for a mastery-avoid goal orientation, the mastery component emerges from optimal antecedents (e.g. motive dispositions, implicit theories, socialization histories) that may facilitate positive consequences (like mastery-approach goals; see Elliot & McGregor, 2001, for a complete discussion). The avoidance component, however, is hypothesized to emerge from non-optimal antecedents and may result in negative consequences (like performance avoidance goals). Conceptually, a mastery-avoid orientation differs from a performance-avoid orientation in terms of the referent or standard used to evaluate performance. Whereas the mastery-avoid referent is relative to the task, the performance-avoid referent is relative to others. Both are similarly valenced, however, whereby the focus is a negative fear of failure. Recent research supports this 2×2 framework of achievement goal orientation (Elliot & McGregor, 2001; Finney, Pieper, & Barron, 2004).

Theoretical models and empirical relations

While theoretical models of achievement goal theory thrive, findings of empirical studies of them and their relations to educationally relevant outcomes are mixed. One possibility (described in more detail below) for these inconsistencies is that they result from limitations in quantifying achievement goals. Although some research has been conducted to examine the psychometric properties of various scales used, we present an alternative method as one possibility to further explore the reliability and validity of common scales used to quantify achievement goal orientation. Before presenting the approach we chose, the two main branches of research on achievement goal theory are reviewed, followed by a description of the Rasch methodology.

As previously noted, reported relations between various learning and achievement variables and a performance-approach goal orientation differ across studies. While some studies showed a performance-approach orientation correlates positively with effective cognitive engagement (e.g. Meece *et al.*, 1988; Wolters, Yu, & Pintrich, 1996) and achievement (Bouffard, Vezeau, & Bordeleau, 1998; Harackiewicz, Barron, Tauer, Carter, & Elliot, 2000; Pintrich, 2000b), some have demonstrated negative relations between performance-approach goals and self-efficacy (e.g. Skaalvik, 1997), as well as positive relations with negative affect and test anxiety (e.g. Kaplan & Maehr, 1999; Meyer, Turner, & Spencer, 1997; Middleton & Midgley, 1997) and with avoidant help seeking (Nadler, 1998; Ryan & Pintrich, 1998). Results are mixed across outcomes, such as beneficial for performance outcomes but detrimental for help seeking. Of more concern are inconsistencies within outcomes. For example, while some studies report positive relations of performance-approach orientation with performance (e.g. Bouffard *et al.*, 1998), others found no relationship between performance-approach goals and performance (e.g. Winne *et al.*, 2006). Similarly, for relations between a performance-approach orientation and self-efficacy, some researchers have found negative relations (e.g. Skaalvik, 1997), while others have found positive relations (e.g. Hinkley, McInerney, & Marsh, 2002) or no relationship (e.g. Winne *et al.*, 2006).

Generally, mastery-approach goals are positively related to learning-related outcomes including cognitive engagement and self-reported self-regulatory strategies (e.g. Pintrich, 2000c), self-efficacy, interest, and value (e.g. Harackiewicz *et al.*, 2000; Pintrich & DeGroot, 1990; Wolters *et al.*, 1996), higher positive affect and lower negative affect (e.g. Kaplan & Maehr, 1999; Middleton & Midgley, 1997), and help seeking (e.g. Ryan &

Pintrich, 1998). Moreover, researchers have found that a mastery-approach orientation is positively related to achievement outcomes (e.g. Church, Elliot, & Gable, 2001; Kaplan & Maehr, 1999). However, not all studies have found consistent positive relations (e.g. Bouffard *et al.*, 1998; Elliot & McGregor, 1999; Ford *et al.*, 1998; Pintrich, 2000b; Wahlstrom, 2001; Winne *et al.*, 2006). Pintrich, for example, found no relationship between a mastery-approach goal orientation and mathematics achievement.

Results have been more consistent with relations between the two avoid orientations and other motivational and learning outcomes. For example, researchers have found negative relationships between a performance-avoid goal orientation and self-efficacy (e.g. Middleton & Midgley, 1997; VandeWalle, Cron, & Slocum, 2001), achievement (e.g. Elliot & McGregor, 1999; VandeWalle *et al.*, 2001; Winne *et al.*, 2006), and various self-reports including metacognition (e.g. Schmidt & Ford, 2003), deep processing strategies (e.g. Elliot & McGregor, 2001; Shih, 2005), intrinsic motivation (e.g. Shih, 2005), help seeking (e.g. Karabenick, 2003), and self-determination (Elliot & McGregor, 2001). Moreover, researchers have also found positive relations between a performance-avoid goal orientation and fear of failure (e.g. Conroy & Elliot, 2004; Elliot & McGregor, 2001), worry, disorganization, test anxiety, and self-reports of surface processing strategies (e.g. Elliot & McGregor, 2001).

Similar patterns have been found between a mastery-avoid goal orientation and various cognitive, motivational, and achievement constructs. Like a performance-avoid goal orientation, researchers have found negative relationships between a mastery-avoid goal orientation and self-determination (e.g. Elliot & McGregor, 2001), help seeking (e.g. Karabenick, 2003), and achievement (e.g. Winne *et al.*, 2006). Positive relationships have also been reported between a mastery-avoid goal orientation and fear of failure (e.g. Conroy & Elliot, 2004; Elliot & McGregor, 2001), worry, disorganization, test anxiety, and self-reported surface processing strategies (e.g. Elliot & McGregor, 2001).

In summary, although various theoretical frameworks define each of the goals similarly, relations among these goals and facets of learning, motivation, and achievement are mixed. Moreover, several studies found no relationship where relations were predicted (e.g. Elliot & McGregor, 1999; Ford *et al.*, 1998; Wahlstrom, 2001; Winne *et al.*, 2006).

Changing goal orientations

Educational psychologists are not solely interested in predicting relations at the group level. Goal theorists also are interested in changing achievement goal orientations and examining those changes within individuals as a function of features of classrooms or school goal structures. They typically found that mastery-approach and performance-approach classrooms and school goal orientations are related to students' adoption of mastery-approach and performance-approach goals (e.g. Nolen & Haladyna, 1990; Roeser, Midgley, & Urdan, 1996; Urdan, 2004; Urdan & Midgley, 2003) as well as students' motivation (e.g. self-efficacy), emotional well-being, cognitive engagement, help seeking, and achievement (e.g. Ames & Archer, 1988; Kaplan & Midgley, 1999; Roeser *et al.*, 1996).

In one study, Urdan and Midgley (2003) examined the effects of changes in students' perceptions of classroom goal structures as they advanced in grade level across elementary to middle school grades. They compared students who perceived an increase, decrease, or no change in mastery-approach and performance-approach classroom goal structures. Students who perceived a mastery-approach increase in their new classrooms also increased their personal mastery-approach goals. Similarly, students who perceived a decrease in mastery-approach goal structures in their new classrooms also decreased their personal mastery-approach goals. Like the mastery-approach increase group, students who perceived an increase in a performance-approach goal structure in their new classrooms also increased their personal performance-approach goals. Moreover, the most negative pattern of changes in motivation, affect, and achievement was associated with a perceived decline in the classroom mastery goal structure.

In another study, Linnenbrink (2005) examined the effects of a quasi-experimentally manipulated classroom goal condition (mastery-approach, performance-approach, and a combined mastery- and performance-approach) and entering personal goal orientations on motivation, help-seeking, emotional well-being, cognitive engagement, and achievement with a sample of upper-year elementary students. In contrast to previous studies, Linnenbrink examined the effects of objective changes in classroom goal structures rather than students' perceptions of those changes. At posttest, Linnenbrink found students in the mastery-approach condition reported higher personal mastery-approach goals than students in the performance-approach condition. Similarly, at posttest, students in the performance-approach condition reported higher personal performance-approach goals than students in the mastery-approach goal condition. No other differences in personal goals were found. Moreover, the classroom goal condition had a statistically detectable effect on achievement and help seeking, with the combined condition demonstrating the greatest benefit.

Personal mastery-approach goals were also beneficial for the majority of outcomes Linnenbrink (2005) examined, including achievement. In contrast, personal performance-approach goals were detrimental for achievement and test anxiety, and unrelated to the remaining outcomes. Surprisingly, the effect of the classroom goal condition did not vary as a function of entering personal goals, which was inconsistent with prior research (e.g. Barron & Harackiewicz, 2001; Harackiewicz & Elliot, 1998). In summary, although research on changing students' goal orientations is not common, studies of this type typically have found that students' personal goal orientations change as a function of the classroom goal structure, although not all studies have found all predicted changes (e.g. Linnenbrink, 2005). Moreover, like research that examined relations between achievement goals and other educationally relevant constructs, results are inconsistent.

Explanations for mixed results

To account for inconsistent results in this field, some researchers have reconsidered the detrimental effects of performance-approach goals (e.g. Harackiewicz *et al.*, 2000; Pintrich, 2000b). The possibility that performance-approach goals may be adaptive has led theorists to suggest that adopting both approach orientations may be the most beneficial. Theorists have labeled this the multiple goals perspective (Barron & Harackiewicz, 2001; Pintrich, 2000a, 2000b). Others, however, claim that a mastery goal perspective is still preferred but propose performance-approach goals may be adaptive for some outcomes and detrimental for others (Kaplan & Middleton, 2002; Midgley *et al.*, 2000). To date, no resolution is established on this issue. Alternatively, other theorists advocate that inconsistencies depend on how achievement is assessed (e.g. Bouffard *et al.*, 1998; Pintrich, 2000b). Unfortunately, these suggestions account for only some inconsistencies, particularly with results pertaining to the performance-approach dimension.

Other possible reasons might explain why relations are inconsistent across studies. For example, one reason might be that researchers use various instruments to quantify achievement goals, and each instrument reflects each of the goal orientations in slightly different ways. This hypothesis is not plausible, however, given that several studies have been conducted to examine the convergent and discriminant validity of several instruments and have found support for similarity of constructs across scales (e.g. Fairchild, Horst, Finney, & Barron, 2005; Jagacinski & Duda, 2001; Midgley *et al.*, 1998; Smith, Duda, Allen, & Hall, 2002).

Alternatively, one could hypothesize the instruments used to quantify achievement goals are not valid and reliable or otherwise not psychometrically sound. Although research has supported the validity of interpretations about constructs and reliability of the data obtained from various goal orientation instruments (e.g. Causgrove, 2000; Conroy, Elliot, & Hofer, 2003; Elliot & McGregor, 2001; Finney *et al.*, 2004; Shih, 2005), assessments of psychometric properties have relied on traditional methods. Specifically, factor analysis has been used to identify the number of goal orientations: two, three, or four orientations. Pearson correlations have been computed for validity, and Cronbach alpha coefficients have been used for reliability. We found no studies that used Rasch analyses to assess psychometric qualities of goal orientation instruments.

The present study

We examine in-depth the psychometric properties of the two most popular achievement goal orientation instruments using both traditional and modern techniques. Our search of literature reveals the two most frequently used instruments are the PALS (Midgley *et al.*, 2000) and the Achievement Goals Questionnaire (AGQ; Elliot & McGregor, 2001) or modest variations of them. As a test case, we scrutinize two aspects of the PALS and AGQ. First, both instruments provide data at the ordinal (ranking) level but not at the interval level. Typically, Likert-scale data are summed and averaged across each item of a subscale to obtain an overall subscale score. These subscale scores are taken to represent a person's score on a particular goal orientation and are used in examining relations to other constructs, such as achievement. We propose that researchers should use Rasch models, or other IRT models, to transform ordinal data into interval data. Since Likert data are regarded as ordinal, it is assumed that the value of each category is higher than the previous category but by an unspecified amount (Bond & Fox, 2001). The Rasch model transforms Likert data into interval scales as logarithmic values of odds (logits). Thus, differences between response choices become mathematically meaningful, as a necessary condition for computing statistics, such as correlations, that assume interval data.

We used aspects of Wilson's (2005) 'four Building-blocks' approach to guide our examination of the psychometric properties of the AGQ and PALS. The four blocks include: (1) the construct map whereby a theoretical construct is defined and mapped to represent more or less of the latent variable, (2) the items design whereby items are constructed to quantify the underlying latent variable, (3) the outcome space in which responses to the items reflect the construct to greater or lesser degrees, and (4) the measurement model which translates scores from the responses back to the construct map to assess how well items reflect the construct. The measurement model is applied to analyse responses to examine model fit, reliability, and validity. Based on results from assessing the model, (a) changes to items, (b) changes to responses, or (c) re-operationalization of the construct may be carried out to improve model fit. We describe how each of the first three facets of Wilson's approach is examined in the context of Rasch modelling next.

Rasch modelling

In the context of Wilson's (2005) framework, a construct is composed of an underlying continuum that can be manifested in two ways - an ordering of the respondents and/or an ordering of the item responses. In classical test theory, it is the ordering of the respondents that is the main focus. In Rasch modelling, however, both are of interest. For a respondent construct map, individuals are ordered from greater to lesser. For this study, for example, individuals can be ordered from more mastery approach oriented to less mastery approach oriented. Alternatively, items can also be ordered from greater to lesser. Item responses may be differentiated to indicate a particular item may be more representative of a high mastery approach than another item. Inherent in the construct, then, is the notion of levels and underlying the construct is a continuum from more to less that can be displayed in a construct map.

According to Wilson (2005), for the outcome space, response options should be well defined, finite and exhaustive, ordered, context-specific, and research based. Of particular interest, for Likert-type scales such as those used for the AGQ and PALS, the order is implicit in the nature of the choices: ***not at all true of me, somewhat true of me, very true of me***. Empirical evidence from Rasch modelling can be used to support the ordering of responses (Wilson, 2005). Finally, for the measurement model, two approaches can be taken: one that focuses on items and their relationship to the construct, and one that focuses on persons' scores and their relationship to the construct. In classical test theory, the focus is on persons' scores; some aggregation of information across items typically based on a summation of item scores. In contrast, the item-focused approach focuses attention on the meaningfulness of the results from the instrument (Wilson, 2005). Wilson's construct modelling approach combines these two approaches together to assess model fit, reliability, and validity.

Accordingly, as Wilson (2005) notes, the Rasch model differs from classical test theory in a number of ways. First, a model is expressed at both the item level and instrument level (e.g. scores). Second, it models the probability of observed responses. Specifically, in the Rasch model, the relationship is the probability of the item response for item I, X_I, modeled as a function of the respondent location θ and the item location δ_i. In achievement and ability contexts, the respondent location is typically labelled ***respondent ability*** and the item location is termed ***item difficulty***. In the context of this study, the locations can be considered as ***attitudes towards achievement goal*** and ***item scale value***, respectively. Thus, each individual has a certain level of each achievement goal orientation (θ) and each item has a certain level of achievement goal orientation it is designed to quantify (δ_i). When the levels are equal, then the individual has a 50% chance of responding positively to the item. When the respondent has more of that construct than the item represents, the probability is greater than .5 that the individual will respond positively. In contrast, when the respondent has less of the construct than the item, the probability is less than .5 that the individual will respond positively.

Compared to traditional psychometric assessments conducted on Likert-type data, a Rasch analysis allows for an examination of two additional facets. These include: whether Likert responses across items are similarly interpreted and similarly ordered, and whether some items are more representative of a construct than other items (e.g. as a function of fit). We investigate the psychometric properties of the AGQ and PALS by further exploring: first, whether each item functions as intended; second, the relative difficulty (item location) of each item along the scale; third, whether each person's responses form a valid response pattern; fourth, each person's relative score (level) on the scale; and, fifth, whether respondent scores and item scores fit together along a similar scale (e.g. similar mean and variance). Finally, we analyse the conceptual structure of the AGQ and PALS by modelling the different types of goal orientations as respondent ability (e.g. level of goal orientation) and item scale value.

STUDY 1

Method

Participants

Participants were 217 undergraduates, one subsample from a Southwestern university (118) in the United States and another from a Northwestern university (99) in Canada. All students were enrolled in an introductory educational psychology class. Thirty-four were male and 181 were female (two did not respond) with an overall mean age of 25.66 ($SD = 8.20$) and a mean self-reported GPA of 3.25 ($SD = 0.48$).

Materials and procedure

We used the original 12-item AGQ (Elliot & McGregor, 2001) to assess students' achievement goals for their educational psychology course. The AGQ is designed to quantify four types of achievement goal orientations: mastery-approach, mastery-avoid, performance-approach, and performance-avoid orientations. At the beginning of the course, students completed the AGQ by indicating the extent to which they judged an item was 'not at all true of me' = 1 to 'very true of me' = 7.

The measurement model

To assess psychometric properties of the AGQ, the Rating Scale model (Andrich, 1978) was implemented using the program Quest (Adams & Koo, 1994). Each item on the AGQ is considered to have six ordered thresholds between response categories corresponding to participants' answers given on a seven-point continuum from 'not at all true of me' to 'very true of me.' Rasch models such as the Rating Scale model yield estimates of respondent ability and item scale score. For rating scales such as the AGQ, respondent ability estimates can be considered to indicate how much an individual espouses a particular goal orientation. The more mastery-approach oriented, for example, the higher that respondent's 'ability' estimate.

To link this back to research on achievement goal orientations, for research that has examined changes in students' goal orientations as a function of classroom environments, respondent ability estimates must be reliable (e.g. like Cronbach's alpha, greater than the guideline of .80). In the context of Rasch modelling, respondent reliability is an index or estimate of the replicability of a respondent's placement that could be expected if the same sample of respondents were given another set of items measuring the same construct (Bond & Fox, 2001).

If respondent ability estimates are not reliable, then changes could be a function of measurement error rather than true changes in people's location on the scale. Similar issues apply for research on relations between achievement goals and other educational variables such as achievement. If respondent abilities are not reliable, relations and effect sizes may be attenuated. Similarly, items must be reliable. Like respondent reliability, item reliability is an estimate of the replicability of item placement within a hierarchy of items along the variable if the same items were given to another comparable sample.

Another issue the Rating Scale model examines has to do with the ordering of item responses along the Likert scale. Disordered item thresholds can be problematic when conventional summated ratings are used (Briggs & Wilson, 2003). In some cases, when the order of the thresholds is not consistent for a set of items, this may not pose a problem when the researcher can explain that result empirically or simply acknowledge that individuals may prefer to respond in that manner. For example, on some instruments, a middle position represents a neutral attitude. It may be the case that respondents find it 'easier' to select a neutral position than to disagree or strongly disagree because they do not want to be disagreeable. If, however, the thresholds were noticeably different from the predictions in the construct map, then that may be a validity issue the researcher should investigate (Wilson, 2005). In other cases, it may not be logical to interpret reversals, such as in cases where it is easier to respond 'strongly agree' versus 'somewhat agree.' In these situations, individuals may select higher scores (e.g. a 5 instead of a 4), which subsequently may influence relations with other constructs. Thus, the purpose of assessing item thresholds is to determine whether the Likert rating scale is ordered as intended.

In the context of the AGQ and PALS, for summation ratings to be meaningful, threshold values must have the same order as the original response options on the original Likert scale. That is, response data must have a hierarchical ordering of responses that correspond to the increasing degree of the latent variable. If threshold order is maintained, and fit and reliability values are within a good range, then a summed scoring approach is warranted but Rasch estimates are still preferred if traditional statistical analyses are conducted that require interval data.

Assessing model fit

To assess model fit, several approaches can be taken. First, one should examine the frequency with which each of the response options is used. As Bond and Fox (2001) note, one should examine shapes of the distributions; normal, bimodal, and slightly skewed distributions are preferable to irregular distributions. Moreover, because response options with low frequencies result in unstable threshold values, the recommended minimal number of responses per category is 10 (Linacre, 1999). Categories with low frequencies or irregular distributions can be remedied by reducing the number of response options, usually by collapsing problematic categories with adjacent ones (Bond & Fox, 2001).

Second, as previously noted, thresholds should increase monotonically; that is, 'somewhat true of me' should have a lower threshold value than 'very true of me,' for example. If thresholds do not increase monotonically, they are disordered. Moreover, the magnitudes of the distances between thresholds should not be too close together or too far apart. Linacre (1999) recommends that thresholds increase by at least 1.40 logits but not more than 5 logits.

Finally, one can examine residuals, that is, the difference between observed scores and expected scores. Fit statistics are weighted and unweighted mean squares that are standardized using the Wilson-Hilferty transformation. Unstandardized mean squares are the mean of the squared residuals and have an expected mean value of 1. Standardized statistics, called infit t and outfit t statistics, have an expected mean of zero and an expected standard deviation of 1. Fit statistics near these mean and standard deviation values suggest a good fit of the data to the measurement model. More specifically, Adams and Khoo (1996) recommend values between 0.75 and 1.33 for mean squares and values between -2 and $+2$ for t are reasonable boundaries to suggest good fit. For mean squares, values greater than 1 suggest more variation in the observed scores than was modelled, and values less than 1 suggest less variation in the observed scores than what was modelled.

Results

Traditional psychometric analyses

Given that two samples were drawn from two different universities, we first explored whether the samples could be pooled. A multivariate analysis of variance was conducted on subscale scores to examine whether group means were equivalent. Box's M test and Levene's test were also computed to examine whether the variance-covariance matrix and variances were equivalent across groups. The multivariate results indicated no difference between groups using Pillai's criterion $F(4, 209) = .126, p = .97$. Box's test of equality of covariances also revealed no difference between groups ($p = .19$), as did Levene's test for equality of variances (all $p > .05$). Accordingly, the two samples were merged into one and all subsequent analyses are based on the combined sample. Data were screened for outliers, normality, linearity, homoscedasticity, multicollinearity, and singularity. All assumptions were met. For example, subscale scores were normally distributed with skewness and kurtosis values within acceptable ranges (e.g. skewness ranged from -0.22 to -0.71, kurtosis ranged from 0.27 to -0.98). Means, standard deviations, and traditional Cronbach alpha reliability coefficients are presented in Table 1. Alpha reliability estimates were good, ranging from .70 to .93.

To test the fit of the four-goal model proposed by Elliot and McGregor (2001), the item data were submitted to a traditional confirmatory factor analysis (CFA). Byrne (1998) suggests that a CFA of an instrument is appropriate when it has been fully developed and its factor structure has been widely validated. The AGQ meets this requirement. We computed the CFA using the EQS software (Bentler & Wu, 1995). CFAs were conducted to assess how well items on the inventory fit the facets the authors identified. Indicators of fit for the CFA were the comparative fit index (CFI) and the root mean squared error of approximation (RMSEA).

To cross-validate Elliot and McGregor's (2001) 2×2 model of goal orientation, we tested five competing models also tested by Elliot and McGregor (2001) and Finney *et al.* (2004): Model A was a four-factor model based on the 2×2 framework for goal orientation. Model B was a three-factor model that included a performance approach factor, a performance avoid factor, and an overall mastery factor formed by summing over mastery approach and mastery avoid items. Model C was a three-factor model that included a mastery approach factor, a performance approach factor, and an overall avoidance factor formed by summing over mastery avoid and performance avoid items. Model D was a two-factor model that included an overall mastery factor and an overall performance factor. Finally, Model E was a two-factor model that included an overall approach factor and an overall avoidance factor. Differences between chi-squared statistics and the CFI were used to compare models. The CFI is particularly sensitive to

Table 1. Descriptive statistics, Cronbach α, and Rasch fit statistics for goal orientations for the AGQ

	Traditional estimates			Rasch item estimates			Rasch person estimates		
Scale	*M*	*SD*	α	*M*	*SD*	α	*M*	*SD*	α
Mastery approach (MAP)	5.77	0.86	.70	0.0	0.54	.97	0.96	0.95	.42
Adjusted mastery approach (AMAP)	5.84	0.74	.72	0.0	1.07	.99	0.56	1.40	.58
Mastery avoid (MAV)	4.25	1.52	.85	0.0	0.29	.92	0.11	1.31	.78
Adjusted mastery avoid (AMAV)	4.36	1.47	.84	0.0	0.36	.92	0.0	1.54	.78
Performance approach (PAP)	4.39	1.60	.93	0.0	0.34	.92	0.60	1.74	.83
Performance avoid (PAV)	4.37	1.57	.76	0.0	0.19	.89	0.20	0.99	.69
Adjusted performance avoid (APAV)	4.47	1.52	.75	0.0	0.23	.88	0.16	1.11	.69

Note. Traditional estimates are reflective of the seven-point scale. Items = 3, $N = 217$ for all subscales. M = Mean, and SD = Standard deviation.

misspecified factor pattern correlations and is a useful fit index when sample sizes are not large (Tabachnick & Fidell, 2001). Values for the CFI greater than or equal to 0.90, and values for the RMSEA less than 0.08 were interpreted as confirming a good fit.

For Model A, the chi-squared, CFI, and RMSEA values were 105.46 ($df = 48, N = 217, p < .01$), 0.95, 0.07, respectively, which suggest a good fit of the data to the model. For model B, the chi-squared, CFI, and RMSEA values were 196.53 ($df = 51, p < .01$), 0.88, and 0.12, suggesting a moderate fit to the data. Chi-square, CFI, and RMSEA values for Model C were 245.22 ($df = 51, p < .01$), 0.84, and 0.13, again suggesting a moderate fit. Model D resulted in a chi-squared of 379.79 ($df = 53, p < .01$), a CFI of 0.73, and an RMSEA of 0.17. Finally, Model E resulted in a chi-squared of 393.39 ($df = 53, p < .01$), a CFI of 0.72, and an RMSEA of 0.17. Consistent with Elliot and McGregor's (2001) and Finney *et al.*'s (2004) results, our data validate the four-factor model of goal orientation for mastery approach, mastery avoidance, performance approach, and performance avoidance.

Rasch model analyses

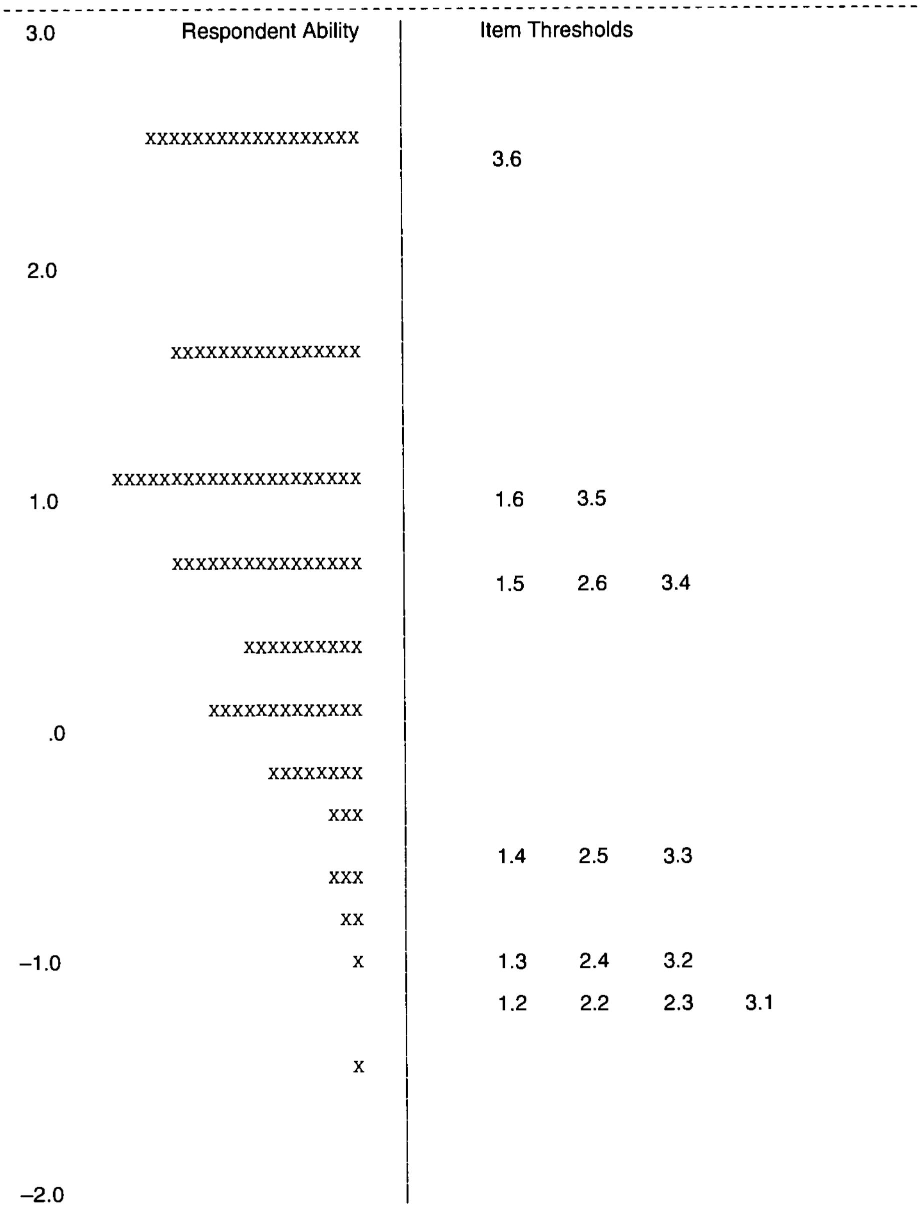
The data were analyzed separately for each of the four goal orientation subscales. Table 1 provides a summary of the traditional and Rasch descriptive statistics and reliability estimates and Table 2 provides Rasch infit and outfit estimates.

Table 2. Fit statistics for subscale scores for the AGQ.

	Infit MS		Outfit MS		Infit *t*		Outfit *t*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Item fit								
MAP	0.78	0.11	0.75	0.06	−2.15	0.99	−1.94	0.28
AMAP	1.00	0.11	0.99	0.13	−0.02	1.17	−0.04	0.99
MAV	1.03	0.32	1.00	0.32	0.06	3.21	−0.19	2.67
AMAV	1.02	0.30	1.01	0.30	0.05	3.02	−0.05	2.50
PAP	0.98	0.10	0.97	0.12	−0.17	0.96	−0.20	0.98
PAV	1.03	0.25	0.96	0.21	0.19	2.52	−0.36	1.75
APAV	1.04	0.25	1.00	0.23	0.27	2.67	−0.02	1.97
Person fit								
MAP	0.73	0.91	0.75	0.95	−0.41	1.09	−0.17	0.88
AMAP	0.97	1.07	0.77	1.19	−0.16	1.20	0.04	0.93
MAV	0.98	1.35	0.99	1.35	−0.27	1.34	−0.11	1.09
AMAV	0.99	1.42	1.00	1.44	−0.27	1.36	−0.10	1.10
PAP	0.97	1.28	0.97	1.27	−0.32	1.38	−0.16	1.13
PAV	0.96	0.96	0.96	0.93	−0.15	1.24	−0.02	0.99
APAV	1.00	1.06	1.00	1.06	−0.20	1.34	−0.05	1.07

Mastery approach

Figure 1 displays the respondent ability distribution (the Xs) and item difficulty thresholds for each of the three items. Values on the left of Figure 1 represent logit values and Xs represent persons. An X for each person is located along the logit scale according to his or her respondent ability estimate. The values on the right of Figure 1 (e.g. 1.1, 1.2, etcetera) represent item thresholds for each rating (e.g. 1, 2, 3, 4, 5, 6,



Note. Each X represents 2 people.

Figure 1. Variable map and distribution of respondent scores for the mastery approach subscale on the AGQ.

and 7) for all three items that quantify a mastery approach. The lead digit in the decimal numbers represents the item number, and the trailing digit represents the value of the response on the Likert scale. For example, 3.1 refers to item #3 and the location represents the probability of response '1' on item 3 is .50. That is, if the respondent is above a particular item threshold, he or she is more than 50% likely to have made that response; if he or she is below a particular item threshold, he or she is less than 50% likely to have made that response.

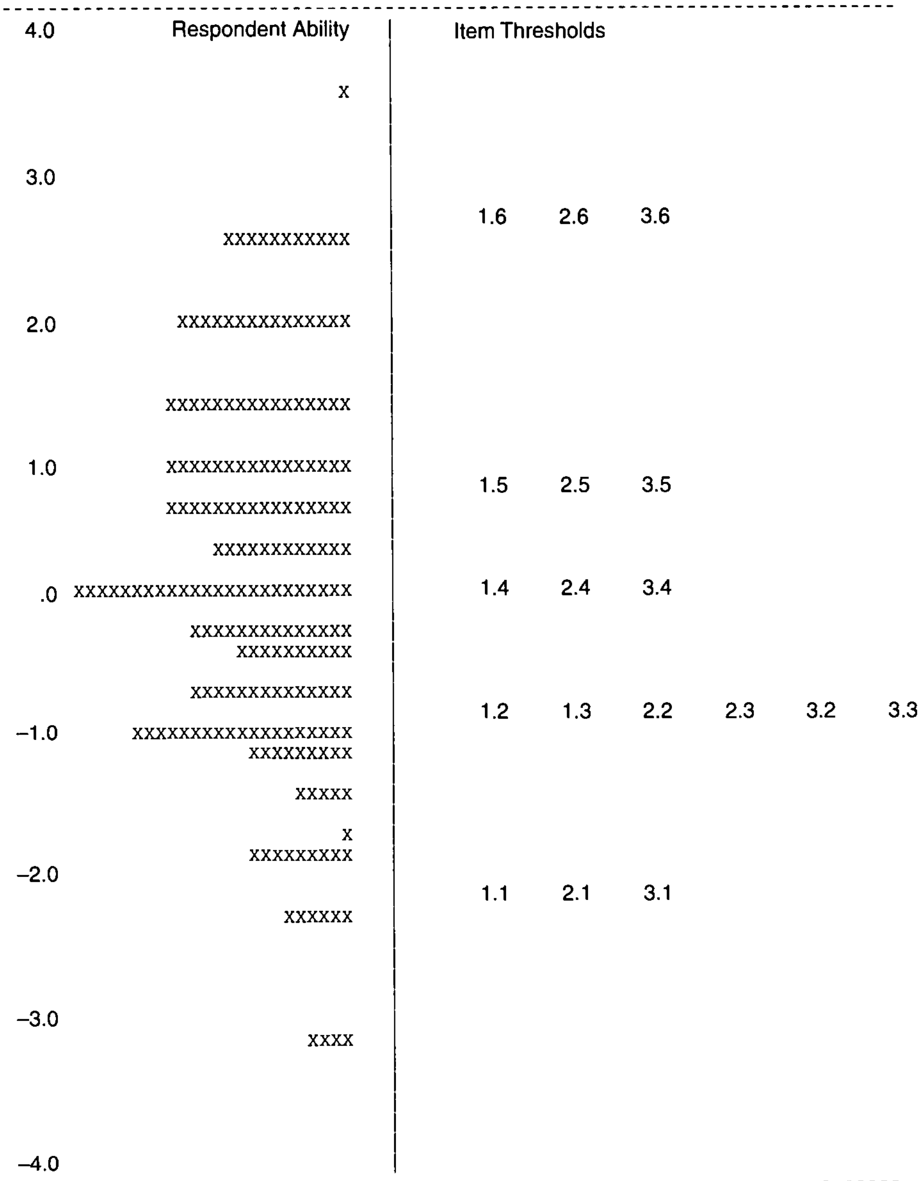
Given the skewed nature of the distribution in Figure 1 and a respondent ability mean of 0.96, our sample found it easy to give high ratings to all three items intended to reflect a mastery approach orientation (a high respondent ability estimate). That is, the average person measure was 0.96 in relation to the average item difficulty of 0.0. The mean should be close to zero for good targeting. Respondent ability estimates ranged from −1.49 logits (the minimum estimate) to 2.52 logits (the maximum estimate). Although respondents found it easy to endorse a mastery approach orientation, the difficulties of the items were well targeted against the respondents and covered the scale's range adequately. Specifically, item thresholds ranged from −1.17 to 2.41 logits (see Figure 1). Moreover, infit and outfit statistics fell within acceptable ranges, specifically, between 0.75 and 1.33 logits for mean square estimates and between −2 and +2 logits for t estimates. This suggests a good fit. However, given that 36 of the respondents had a higher logit value than 2.41, more items need to be included that are more difficult to endorse positively. In other words, this scale has a ceiling effect that limits respondents from expressing the strength of their mastery approach goal orientation.

As can be seen in Table 1, the respondent reliability estimate for respondent ability was low (.42), whereas item reliability was high (.97). By examining the distribution of responses, it is evident that no participant selected '1' or '3' as an option on the statement, 'It is important for me to understand the content of this course as thoroughly as possible.' Also, for the statement, 'I want to learn as much as possible from this class,' no one responded using '1.' Moreover, for all three items, frequencies were less than 10 for the first 3 response options of 1, 2, and 3. Additionally, as shown in Figure 1, item thresholds were not at least 1.40 logits apart. Specifically, a disordered increase in categorical ratings resulted for the item #2 'It is important for me to understand the content of this course as thoroughly as possible.' For example, for this statement, the ratings ranged in difficulty from 3 and 4, 5, 6, and 7, rather than from 1, 2, 3, 4, 5, 6, and 7. That is, individuals did not distinguish between responses 3 and 4 (2.2 and 2.3 in Figure 1). These results indicate the mastery approach items did not differentiate more mastery approach from less mastery approach oriented students because most students had high ratings on these items.

As Bond and Fox (2001) suggest, when issues arise with threshold disorders or frequencies of less than 10 responses for a response option, adjacent response options should be merged and fit indices recalculated to assess whether model fit improves. As a test case, we ran additional analyses to examine this possibility. Item response options 1, 2, and 3 were merged into response option 4. As shown in Tables 1 and 2, results for the adjusted mastery approach subscale revealed a better fit to the model. Respondent separation reliability increased from .42 to .58, and item separation reliability increased from .97 to .99. Unstandardized and standardized fit statistics also improved. Although fit improved, differentiating individuals along the mastery approach subscale is still problematic given the unsatisfactory reliability estimates obtained for the adjusted scale.

Mastery avoid

Figure 2 displays the respondent ability distribution and item difficulty thresholds for each of the three items. As shown in Table 1, the near-zero mean (0.11) for the mastery avoid subscale suggests the items were well targeted to our sample's perceptions of mastery avoid (Bond & Fox, 2001). Respondent ability estimates ranged from −3.06 to 2.61 logits and item difficulties ranged from −2.06 to 2.70 logits. Given that only 10 students had logit values less than −2.06, these results show that students were well



Note. Each X represents 1 person.

Figure 2. Variable map and distribution of respondent scores for the mastery avoid subscale on the AGQ.

targeted by the mastery avoid subscale. Moreover, infit and outfit statistics were within acceptable ranges suggesting a good fit. However, standard deviations were large which suggests items do not satisfactorily differentiate respondent ability for individuals similar on the mastery avoid scale.

The reliability estimate for respondent ability was good (.78) but less than the reliability for the item estimates (.92). Like the mastery approach subscale, item thresholds for mastery avoid shown in Figure 2 did not progress in an ordered fashion for all three items nor were distances between all threshold 1.40 logits or more. Responses of 3 or 4 were not distinguishably different; that is, a rating of 3 was interpreted by our participants the same as 4. Like the mastery approach subscale, we merged categories to examine whether fit improved; response options 3 and 4 were combined. As shown in Tables 1 and 2, merging the two categories did not result in a better fit. Reliability estimates remained the same and traditional reliability slightly decreased.

Performance approach

Figure 3 displays the respondent ability distribution and item difficulty thresholds for each of the three items. Based on the respondent ability mean of 0.60, students in our sample found it easy to give more difficult (e.g. higher) ratings to items on the scale. Respondent ability estimates ranged from − 3.90 to 4.17 logits and item thresholds ranged from − 3.01 to 3.31 logits. The majority of the respondents were with the range of − 3.00 to 3.16 logits, however, which suggests the students were well targeted by the scale. Moreover, infit and outfit statistics fell within the acceptable ranges suggesting a good fit.

As shown in Table 1, the reliability estimate for respondent ability was good (.83) but less than the reliability for item estimates (.92). Moreover, item thresholds also progressed in an ordered fashion. These results demonstrate that, in general, items on the performance approach scale have good psychometric properties.

Performance avoid

Figure 4 displays the respondent ability distribution and item difficulty thresholds for each of the three items. The near-zero mean (0.20, Table 1) for the performance-avoid subscale suggests that the items were well targeted to our sample's ability. Respondent ability estimates ranged from − 2.35 to 2.43 logits and item thresholds ranged from − 1.34 to 1.39 logits. These results indicate that students in the extreme ends of the distributions are not well targeted by the scale. Students with a low or high performance avoid orientation are not well differentiated. Item thresholds tend to cluster around zero.

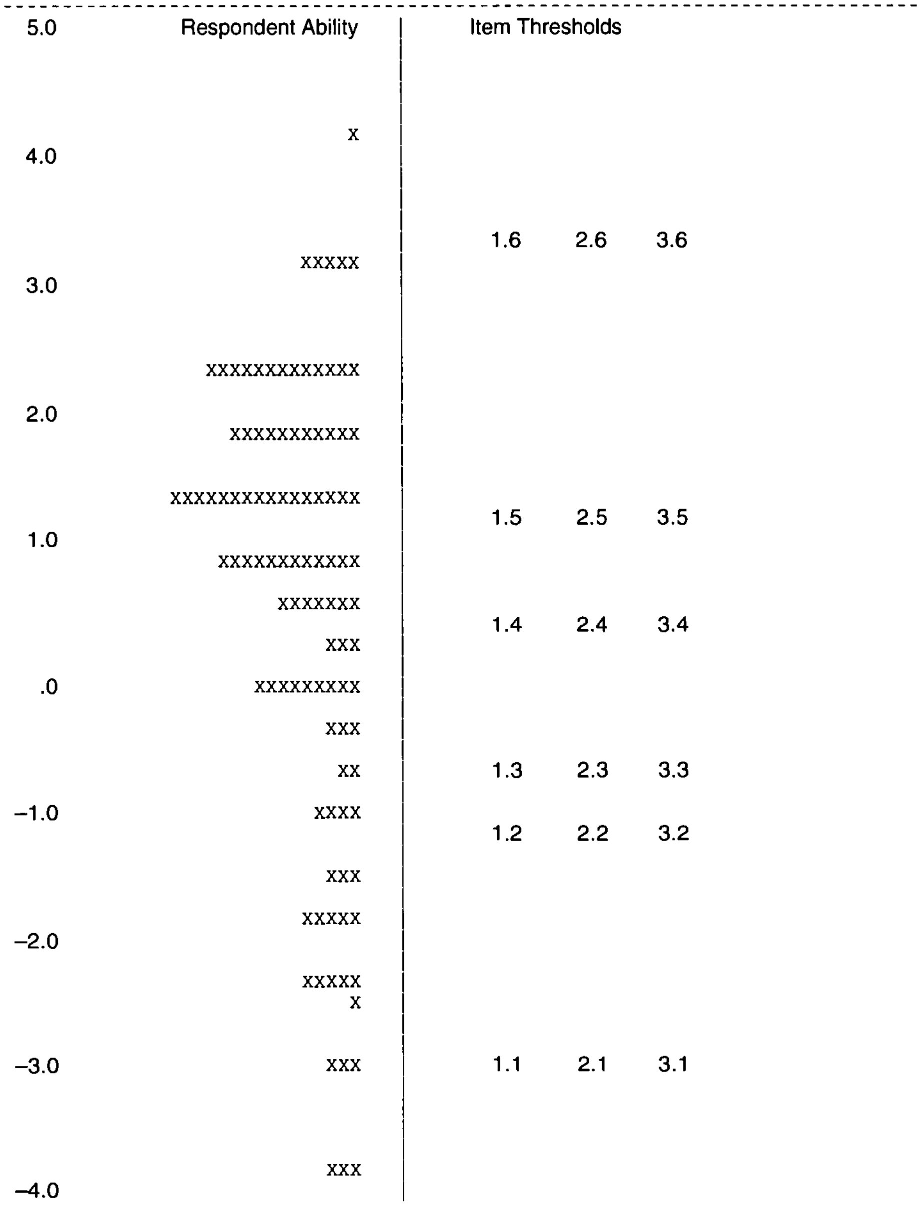
The reliability estimate for respondent ability was moderate (.69) whereas reliability for item estimates was good (.89). As shown in Figure 4, item thresholds were not at least 1.40 logits apart. A disordered increase in categorical ratings resulted for all three items. All three resulted in a reversal of categories 3 and 4. Unstandardized and standardized infit and outfit statistics (Table 2) were within acceptable ranges but large variability in standardized infit and outfit statistics suggests performance avoid items did not differentiate individuals with similar scores, particularly at the lower and upper ends of the distribution, as previously noted. To examine whether merging adjacent categories resulted in improved fit, we recomputed the analyses with response categories 3 and 4 combined. As shown in Tables 1 and 2, fit did not improve; reliability estimates were equivalent.

STUDY 2

Method

Participants

A different sample of 126 undergraduates from a Southwestern university in the United States participated in the second study. All students were enrolled in an educational



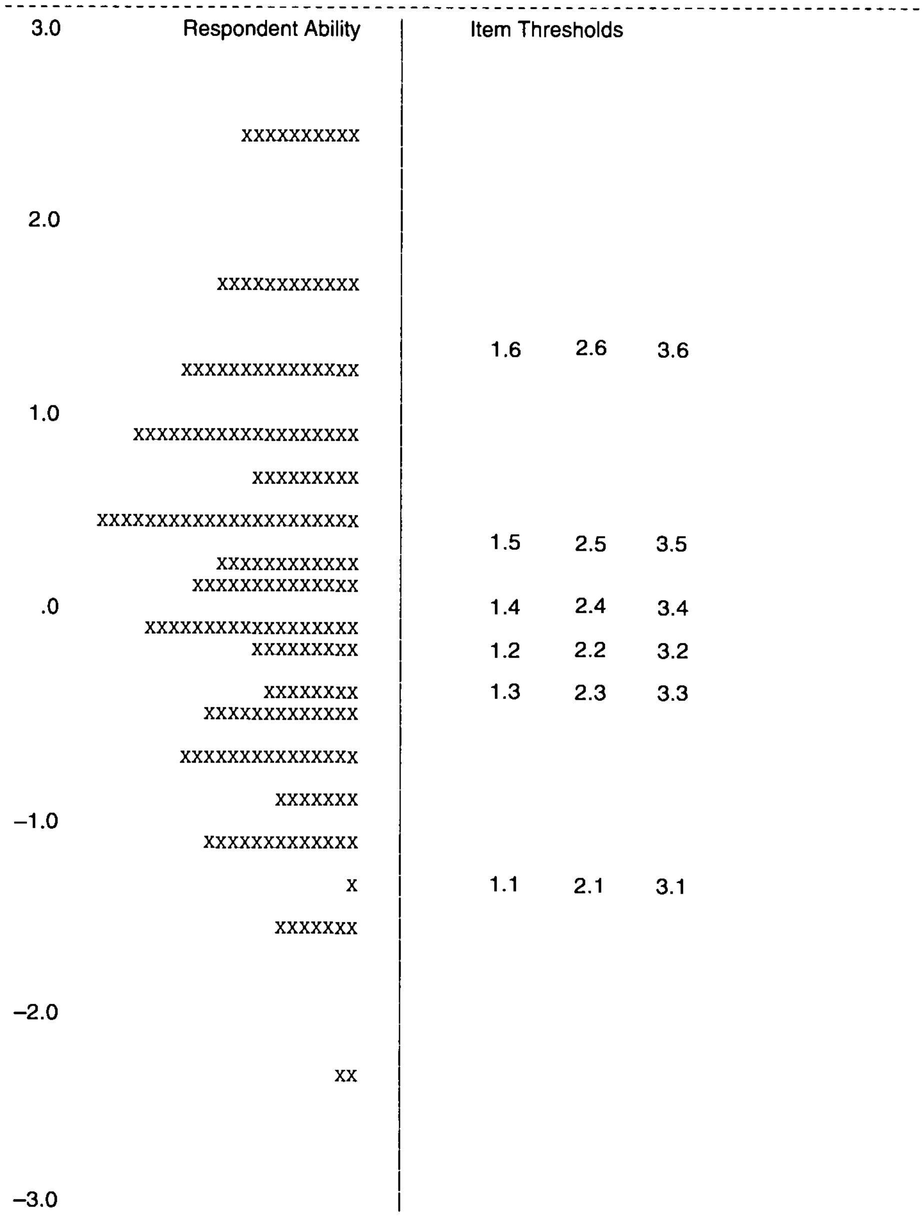
Note. Each X represents 2 people.

Figure 3. Variable map and distribution of respondent scores for the performance approach subscale on the AGQ.

psychology class. Thirty were male and 95 were female (one did not respond) with an overall mean age of 24.63 ($SD = 7.16$) and a mean self-reported GPA of 3.25 ($SD = 0.42$).

Materials and procedure

We used the original 14-item PALS (Midgley *et al.*, 2000) to assess students' achievement goals for their educational psychology course. Students completed the PALS by



Note. Each X represents 1 person.

Figure 4. Variable map and distribution of respondent scores for the performance avoid subscale on the AGQ.

indicating the extent to which they judged an item was 'not at all true' = 1, 'somewhat true' = 3, or 'very true' = 5 of them (ratings of 2 and 4 did not have an anchor). As was done in all prior studies, our participants' responses were then averaged across items to form the three goal orientations; mastery approach (5 items), performance approach (5 items), and performance avoid (4 items).

Measurement model

The same measurement model was used in Study 2 as in Study 1.

Results

Traditional psychometric analyses

Data were screened for outliers, normality, linearity, homoscedasticity, multicollinearity, and singularity. All assumptions were met. For example, all subscale scores computed using traditional sums to responses were normally distributed with skewness and kurtosis values within acceptable ranges: skewness ranged from -0.57 to 0.43, kurtosis ranged from -0.64 to -0.37. Means, standard deviations, and traditional Cronbach alpha reliability coefficients are presented in Table 3. Alpha reliability estimates were good, ranging from .84 to .92.

Table 3. Descriptive statistics, Cronbach α, and Rasch fit statistics for goal orientations for the PALS.

	Traditional estimates			Rasch item estimates			Rasch Person estimates		
Scale	*M*	*SD*	α	*M*	*SD*	α	*M*	*SD*	α
Mastery approach	4.19	0.70	.92	0.0	0.77	.96	1.14	1.13	0.64
Adjusted mastery approach	4.23	0.64	.92	0.0	1.07	.94	0.29	2.14	0.78
Performance approach	2.51	0.99	.90	0.0	0.78	.96	−0.90	1.83	0.85
Performance avoid	2.65	1.00	.84	0.0	0.37	.94	−0.29	0.78	0.56

Note. Traditional estimates are reflective of the five-point scale. Items = 5, $N = 126$ for all subscales. M = Mean, and SD = Standard deviation.

To test the fit of the three-goal model proposed by Midgley *et al.* (2000), the data were submitted to a traditional confirmatory factor analysis (CFA). Like Study 1, we computed the CFA using the EQS software (Bentler & Wu, 1995). CFAs were conducted to assess how well items on the inventory fit the goal orientations the authors identified. Indicators of fit for the CFA were the comparative fit index (CFI) and the root mean squared error of approximation (RMSEA).

To cross-validate Midgley et al.'s (2000) model of goal orientation, we tested three competing models: Model A was a three-factor model based on their framework for goal orientation. Model B was a two-factor model that included an overall performance factor formed by summing over performance approach and performance avoid items. Model C was a one-factor model where all items were combined on one scale.

For Model A, the chi-squared, CFI, and RMSEA values were 190.39 ($df = 74$, $N = 126$, $p < .01$), .92, .11, respectively, which suggest a good fit of the data to the model. For model B, the chi-squared, CFI, and RMSEA values were 204.50 ($df = 76$, $p < .01$), .90, and .12, also suggesting a good fit to the data. These results indicate the two performance subscales are correlated. Because the models were close in fit, we compared these two models to examine whether the more complex three-factor model resulted in a better fit. Using the likelihood ratio test, results revealed the more complex model was a better fit ($p < .01$). Finally, chi-squared, CFI, and RMSEA values for Model C were 639.15 ($df = 77$, $p < .01$), .56, and .24, suggesting a poor fit. Consistent with previous studies (e.g. Middleton & Midgley, 1997; VandeWalle, 1997), our data validate the three-goal model of goal orientation for mastery approach, performance approach, and performance avoidance. Although the two performance subscales were correlated, modelling them as distinct latent factors resulted in the best fit, which provides further empirical support for Midgley et al.'s (2000) theoretical framework.

Rasch model analyses

The data were analyzed separately for each of the three goal orientation subscales. Table 3 provides a summary of the traditional and Rasch descriptive statistics and reliability estimates and Table 4 provides Rasch infit and outfit estimates.

Table 4. Fit statistics for subscale scores for the PALS.

	Infit MS		Outfit MS		Infit *t*		Outfit *t*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Item fit								
Mastery approach	0.43	0.07	0.47	0.07	−4.93	0.81	−3.40	0.46
Adjusted mastery approach	0.99	0.12	0.97	0.17	0.0	0.75	−0.09	0.69
Performance approach	1.02	0.20	0.99	0.22	0.13	1.42	−0.04	1.23
Performance avoid	0.48	0.11	0.49	0.08	−5.27	1.47	−3.57	0.86
Person fit								
Mastery approach	0.73	0.91	0.75	0.95	−0.41	1.09	−0.17	0.88
Adjusted mastery approach	0.98	0.85	0.97	1.03	−0.11	1.13	0.01	0.98
Performance approach	1.02	1.05	0.99	1.04	−0.17	1.41	−0.07	1.14
Performance avoid	0.49	0.39	0.49	0.39	−0.92	1.00	−0.55	0.71

Mastery approach

Figure 5 displays the respondent ability distribution and item difficulty thresholds for each of the five items. As shown in Table 3, the reliability estimate for respondent ability was low (0.64), whereas item difficulty reliability was high (0.96). Additionally, some item threshold distances were not greater than 1.40. Specifically, as shown in Figure 5, a disordered increase in categorical ratings resulted for item #1 'It's important to me that I learn a lot of new concepts this year,' item #3 'One of my goals is to master a lot of new skills this year,' item #4 'It's important to me that I thoroughly understand my class work,' and item #5 'It's important to me that I improve my skills this year.' For these statements, the ratings ranged in difficulty from 2, 4, 3, and 5, rather than from 1, 2, 3, 4, and 5. The statement, 'One of my goals in class is to learn as much as I can,' also had disordered values, ranging from 1, 3, 2, 4, and 5.

As shown in Table 4, unstandardized infit and outfit statistics were not within the acceptable range (0.43 and 0.47, respectively) For example, the infit value of 0.43 indicates an overfit of the data to the Rasch model. That is, there was 57% less variation in the observed data than what the Rasch model predicted, which suggests the response pattern was too determined (Bond & Fox, 2001). The standardized infit and outfit t statistics also revealed a poor fit to the model (values were −4.93 and −3.40, respectively), again suggesting overfit and too little variation. Respondent ability estimates ranged from −1.27 logits to 3.30 logits, whereas item difficulty thresholds ranged from −.35 to 2.20, which also suggest a poor fit. Like the AGQ, the majority of

```
------------------------------------------------------------------------------
4.0        Respondent Ability  |   Item Thresholds
                               |
                   XXXXXXXXXX  |
3.0                            |
                               |
                    XXXXXXXXX  |
                               |   2.4
2.0                            |
                      XXXXXXX  |
                               |   1.4    2.3    3.4    4.4    5.4
                    XXXXXXXXX  |
1.0       XXXXXXXXXXXXXXXXXXX  |
               XXXXXXXXXXXXXX  |
                       XXXXXX  |
                       XXXXXX  |
.0                      XXXXX  |
                          XXX  |   1.2    2.1    3.2    4.2    5.2
                            X  |   1.3    2.2    3.3    4.3    5.3
                           XX  |
-1.0                           |
                          XXX  |
                               |
-2.0                           |
------------------------------------------------------------------------------
```

Note. Each X represents 1 person.

Figure 5. Variable map and distribution of respondent scores for the mastery approach subscale on the PALS.

individuals chose higher values on the Likert scale (e.g. 4 and 5), with very few individuals selecting lower values. For four of the five items, no one selected a value of 1. Finally, based on the respondent ability mean of 1.14, our sample found it easy to give more difficult (e.g. higher) ratings to items on the scale. These results suggest the mastery approach items did not differentiate students with moderate mastery approach from less mastery approach oriented students, which is also apparent given the low reliability estimate.

Like Study 1, we performed additional analyses to examine whether merging adjacent response categories improved meaning and measurement function. Item response options 1 and 2 were merged into response option 3. As shown in Tables 3 and 4, results for the adjusted mastery approach subscale revealed a better fit to the model for the person estimates but not for the traditional or item estimates. Rather, reliability estimates stayed the same or slightly decreased. Unstandardized and standardized fit statistics improved, however.

Performance approach

Figure 6 displays the respondent ability distribution and item difficulty thresholds for each of the five items. The reliability estimate for respondent ability was good (.85), but less than the reliability for the item estimates (.96). As shown in Figure 6, item thresholds progressed in an ordered fashion and distances between thresholds were at least 1.40 logits or more. Unstandardized and standardized infit and outfit statistics (Table 4) were near expected values, which suggests a good fit. The respondent ability estimates ranged from −4.32 to 4.21, whereas item thresholds ranged from −2.68 to 2.57. Item difficulties were not well targeted against the respondents and did not adequately cover the range. Based on the respondent ability mean of −0.90 (Table 3), students in our sample found it difficult to give more difficult (e.g. higher) ratings to items on the scale. These results suggest that, in general, items on the performance approach scale have good psychometric properties.

Performance avoid

Figure 7 displays the respondent ability distribution and item difficulty thresholds for each of the five items. The mean (−0.29, Table 3) for the performance-avoid subscale suggests that the items were a little more difficult than our sample's ability. The reliability estimate for respondent ability was poor (0.56), whereas reliability for items was good (0.94). As shown in Figure 7, item thresholds were not consistent. A disorered increase in categorical ratings resulted for item #1, 'It's important to me that I don't look stupid in class.' For this item, categories 4 and 5 were not distinguishable (e.g. threshold values were the same). Moreover, little differentiation was made in difficulty between ratings 2, 3, 4, and 5 and thresholds were not more than 1.40 logits apart.

Unstandardized infit and outfit statistics were not within the acceptable ranges (0.48 and 0.49, respectively, Table 4). Like the mastery approach subscale, there was an overfit of the data to the model. The infit value of 0.48 indicates there was 52% less variation in the observed data than what the Rasch model predicted, which suggests the response pattern was too determined (Bond & Fox, 2001). Standardized infit and outfit t statistics also revealed a poor fit to the model (values were −5.27 and −3.57, respectively), with values less than zero indicating too little variance. Moreover, respondent ability range was from −1.91 to 1.96, whereas item thresholds ranged from −0.26 to 0.49. This suggests the difficulties of the items were not well targeted to our sample and did not adequately cover the range. The majority of item thresholds centred around zero. Thus, items on the PALS did not satisfactorily distinguish between individuals low on performance avoid versus high on performance avoid. Given how close item thresholds were for this subscale, there was no clear strategy for improving fit. Alternate solutions, as Wilson (2005) recommends, are to rewrite items or include more items to better differentiate respondents along this subscale.

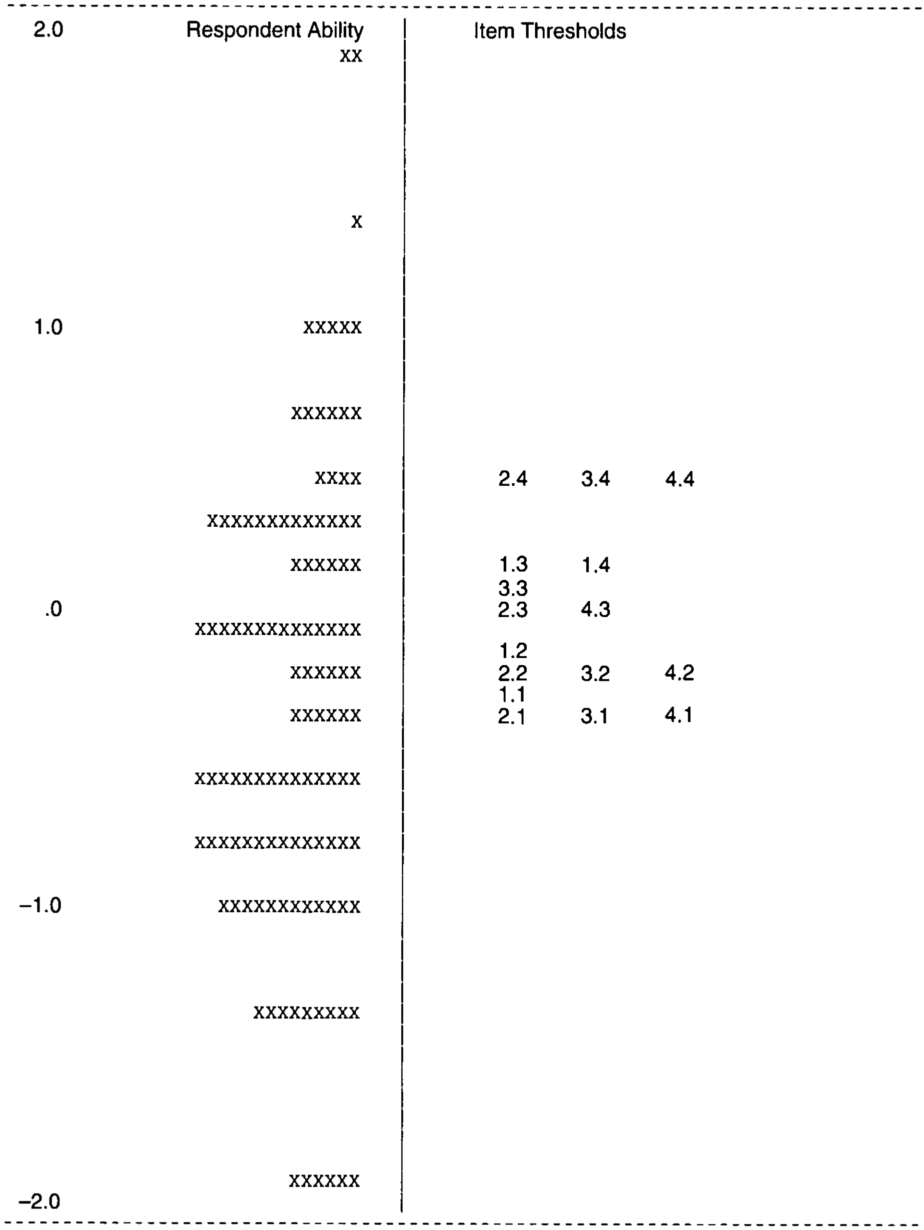
```
-------------------------------------------------------------------------
 5.0          Respondent Ability  |  Item Thresholds
                                  |
                                X |
 4.0                              |
                                  |
                                X |
                                  |
 3.0                              |
                                  |  1.4   2.4   3.4   4.4   5.4
                              XXX |
 2.0                              |
                            XXXXX |
                             XXXX |
 1.0                        XXXXX |
                                  |  1.3   2.3   3.3   4.3   5.3
                              XXX |
                       XXXXXXXXXX |
  .0                      XXXXXXX |
                      XXXXXXXXXXX |
                          XXXXXXX |  1.2   2.2   3.2   4.2   5.2
-1.0                      XXXXXXX |
                      XXXXXXXXXXX |
                   XXXXXXXXXXXXXX |
-2.0                              |
                             XXXX |
                                  |  1.1   2.1   3.1   4.1   5.1
                           XXXXXX |
-3.0                              |
                           XXXXXX |
                                  |
-4.0                              |
                       XXXXXXXXXX |
                                  |
-5.0                              |
-------------------------------------------------------------------------
```

Note. Each X represents 1 person.

Figure 6. Variable map and distribution of respondent scores for the performance approach subscale on the PALS.

Discussion

We examined, the psychometric properties of the Achievement Goal Questionnaire (Elliot & McGregor, 2001) and the Patterns of Adaptive Learning Scale (Midgley *et al.*, 2000), using traditional and more modern statistical techniques, to clarify some inconsistencies across studies and to further examine the reliability and evidence for validity. Results of traditional psychometric assessments of the AGQ and PALS replicated



Note. Each X represents 1 person.

Figure 7. Variable map and distribution of respondent scores for the performance avoid subscale on the PALS.

previous studies. Reliability estimates ranged from good to very good for raw subscale scores. Fit indices from confirmatory factor analyses confirmed the goodness of fit for each hypothesized model. Based on traditional assessments, the AGQ and PALS are accurate indicators of achievement goals.

Results from modern psychometric analyses revealed a number of issues with the majority of scales that need to be addressed in future research. First, Rasch analyses revealed that estimates of reliability for the items were very good, whereas respondent ability estimates varied from poor to good. This finding is important because it indicates that although items can reliably reflect a group's aggregate goal orientation, using either instrument to characterize an individual's goal orientation is hazardous. For the two approach subscales on the AGQ and the mastery approach subscale on the PALS, individuals did not select the lower portion of the response scale. This indicates that individuals are more likely to highly agree to items on the approach subscales, which limits variation in responses. We suggest the AGQ and PALS be revised to include more cleverly worded items that include approach orientations that individuals are less likely to endorse. Increasing the number of items per subscale may provide better segmenting of person samples, thereby increasing meaningfulness of comparisons. That is, increasing the range of item difficulty will improve variable definition and will increase the precision of person estimates.

Operationally, items with which fewer respondents agree are needed to differentiate between less approach-oriented and more approach-oriented students. Also, it should be noted that when students typically give higher ratings on these two subscales, variance is decreased and will attenuate correlations with other constructs such as achievement. This might help to explain why some researchers did not find statistically detectable relationships between the two approach orientations and other constructs.

Two other issues we encountered were disordered patterns of responses on the mastery-approach, mastery-avoid, and performance-avoid goal orientation subscales on the AGQ, and mastery approach and performance avoid subscales on the PALS. These results suggest that, when examining changes in individuals' goal orientations, a traditional approach to computing subscale scores by summing responses to items is not warranted with either instrument. As previously noted, Likert scales provide ordinal estimates. Accordingly, Rasch modelling of data satisfies statistical analyses that assume interval data.

Based on these results, we further suggest the number of category responses may need reconsideration. It may be that five categories, rather than seven, would be more appropriate for the AGQ. If individuals have only 5 response options from which to choose and each option is anchored with a descriptive label, then confusions may be less likely. As Fowler (2002) recommends, anchoring every number with a descriptor provides more interpretive consistency across respondents. In contrast, when numbers are not labelled, individuals are left to interpret what numbers represent and individuals' interpretations may vary which may then increase measurement error. Accordingly, improving item format may help further our understanding of the nature of achievement goal orientation and its relations to other constructs.

In summary, when a traditional approach to forming subscales is compared to a modern Rasch model approach, the construct as represented by items is well quantified on all goal orientations. Evidence for this is that reliability estimates were high for our sample. Thus, when researchers examine relations between goal orientations and various outcomes, it is unlikely that conflicting results we found in our review of the literature were a function of measurement error from individual items. Rather, it could be the case the differences found between studies were a function of the samples drawn, variations in sample sizes across studies, and the different instruments that were used to quantify achievement goals. We note, however, that for the PALS the two performance subscales were correlated. Given that the two-factor model resulted in a good fit, this suggests the performance approach and performance avoid goal orientation subscales are not behaving as the theoretical model assumes. Interestingly, the one goal orientation that resulted in the best fit and highest reliability estimates was the performance approach subscale for both the AGQ and PALS, but with different samples. Future research is needed whereby individuals respond to both the AGQ and PALS to examine, whether the two instruments quantify the same constructs and whether individuals are similarly targeted for each of the goal orientations.

Given the majority of studies we reviewed revealed the performance-approach subscale as the most conflicting with respect to relations to other educational variables, we posit that contradictory results may not be a function of measurement issues with the performance approach subscales. Rather, contradictory evidence may be a result of broader theoretical issues. Although theoretical models of achievement goal orientation have developed from two-dimensional to three- or four-dimensional models, it appears more work is needed to create items that are better at distinguishing these two types of orientations. As Wilson (2005) suggests, when data do not function as intended, items, responses to those items, or the construct map itself should be reconsidered.

Given the low reliability estimates for respondent ability, individuals are not reliably targeted on the achievement goal subscales. Accordingly, we recommend that when researchers are interested in examining changes in individuals' goal orientations over time or across contexts, a Rasch approach is warranted. Because person-centred approaches in the achievement goal literature are increasing in frequency, this has serious implications for this line of research. If respondent reliability is low, empirical evidence may continue to produce conflicting results. Advancing theoretical specifications of models based on empirical results with moderate to high measurement error is hazardous.

In conclusion, tracking individuals across contexts and time is critical for advancing achievement goal theory. Our results have several implications in this respect. First, better scale development for the widely used PALS and AGQ is warranted. Although traditional psychometric assessments of both scales have demonstrated that both are valid and reliable, results of our Rasch analyses suggest that, while item difficulties are reliable estimates (e.g. one would expect similar item difficulties on subsequent tests on the same sample), respondent ability estimates are not. That is, when individuals complete the instruments on more than one occasion, as they must be in research on changing goal orientations, changes in traditionally formed scores from pre-test to posttest may not be reliable reflections of changes in goal orientations. Accordingly, one must question some conclusions that have been offered based on research that examined the influence of contextual factors on goal orientations. To what extent do contextual factors influence goal orientations? Are goal orientations more or less stable than researchers have postulated (e.g. Ames, 1992; Dweck & Leggett, 1988)?

Finally, mastery approach subscales in both the AGQ and PALS failed to satisfactorily fit their respective model, and, for both instruments, distributions for mastery goals were negatively skewed. Moreover, statistically detectable positive correlations between subscales were found, and this was particularly high between the performance subscales on the PALS. These results invite questions about the various theoretical models that have been proposed. Given moderate correlations between the two avoidance goals, are they theoretically that different? Do individuals adopt only one goal type or are multiple goals more common? As well, the debate has yet to be resolved about whether the normative perspective, that a mastery approach goal leads to optimal outcomes, or the multiple goals perspective, both a mastery approach and performance approach adoption leads to optimal outcomes, is preferred. Without improving the scales currently used to quantify achievement goals, these theoretically pressing questions may remain unresolved.

Acknowledgements

Support for this research was provided by grants to Philip H. Winne from the Social Sciences and Humanities Research Council of Canada (410-2002-1787 and 512-2003-1012), the Canada Research Chair program, and Simon Fraser University.

References

Adams, R. J., & Khoo, S. T. (1994). *Quest: The interactive test analysis system*. Camberwell, Victoria: ACER.

Adams, R. J., & Khoo, S. T. (1996). *Quest: The interactive test analysis system*. Melbourne, Australia: ACER.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261-271.

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*, 260-270.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561-573.

Barron, K. E., & Harackiewicz, J. M. (2001). Achievement goals and optimal motivation: Testing multiple goal models. *Journal of Personality and Social Psychology, 80*, 706-722.

Bentler, P. M., & Wu, E. J. C. (1995). *EQS for windows: User's guide*. Encino, CA: Multivariate Software, Inc.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bouffard, T., Vezeau, C., & Bordeleau, L. (1998). A developmental study of the relation between combined learning and performance goals and students' self-regulated learning. *Journal of Educational Psychology, 68*, 309-319.

Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement, 4*, 87-100.

Butler, R. (1993). Effects of task- and ego-achievement goals on information seeking during task engagement. *Journal of Personality and Social Psychology, 65*, 18-31.

Byrne, B. M. (1998). *Structural equation modelling with LISREL, PRELIS, and SIMPLIS*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Causgrove, D. J. (2000). Goal orientations, perceptions of the motivational climate, and perceived competence of children with movement difficulties. *Adapted Physical Activity Quarterly, 17*, 1-19.

Church, M. A., Elliot, A. J., & Gable, S. L. (2001). Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology, 93*, 43-54.

Conroy, D. E., & Elliot, A. J. (2004). Fear of failure and achievement goals in sport: Addressing the issue of the chicken and the egg. *Anxiety, stress and coping: An International Journal, 17*, 271-285.

Conroy, D. E., Elliot, A. J., & Hofer, S. M. (2003). A 2 × 2 achievement goals questionnaire for sport: Evidence for factorial invariance, temporal stability, and external validity. *Journal of Sport and Exercise Psychology, 25*, 456-476.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*, 256-273.

Edwards, O., & Muis, K. R. (2008). The stability of achievement goal orientations. Paper presented at the American Educational Research Association, New York.

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist, 34*, 169-189.

Elliot, A. J., & Church, M. A. (1997). A hierarchical model to approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 72*, 218-232.

Elliot, A. J., & Harackiewicz, M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology, 70*, 461-475.

Elliot, A. J., & McGregor, H. (1999). Test anxiety and the hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 76*, 628-644.

Elliot, A. J., & McGregor, H. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*, 501-519.

Fairchild, A. J., Horst, S. J., Finney, S. J., & Barron, K. E. (2005). Evaluating existing and new validity evidence for the academic motivation scale. *Contemporary Educational Psychology, 30*, 331-358.

Finney, S. J., Pieper, S. L., & Barron, K. E. (2004). Examining the psychometric properties of the achievement goal questionnaire in a general academic context. *Educational and Psychological Measurement, 64*, 365-382.

Fisher, S. L., & Ford, J. K. (1998). Differential effects of learner effort and goal orientation on two learning outcomes. *Personnel Psychology, 51*, 397-420.

Ford, J. K., Smith, E. M., Weissbein, D. A., Gully, S. M., & Salas, E. (1998). Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer. *Journal of Applied Psychology, 83*, 218-233.

Fowler, F. J., Jr (2002). *Survey research methods* (3rd ed.). Newbury Park, CA: Sage.

Fryer, J. W., & Elliott, A. J. (2007). Stability and change in achievement goals. *Journal of Educational Psychology, 99*, 700-714.

Harackiewicz, J. M., Barron, K. E., Tauer, J. M., Carter, S. M., & Elliot, A. J. (2000). Short-term and long-term consequences of achievement goals: Predicting interest and performance over time. *Journal of Educational Psychology, 92*, 316-330.

Harackiewicz, J. M., & Elliot, A. J. (1998). The joint effects of target and purpose goals on intrinsic motivation: A mediational analysis. *Personality and Social Psychology Bulletin, 24*, 675-689.

Hinkley, J. W., McInnerney, D. M., & Marsh, H. W. (2002). Positive and negative ability beliefs among Navajo high school students: How do they relate to students' school achievement goals? *Paper presented at the Annual Meeting of the American Educational Research Association*, New Orleans.

Hoover, P. B., Steele-Johnson, D., Beauregard, R. S., & Schmidt, A. M. (1999, April). Learning and performance goal orientation interactions with dynamic task complexity. *Paper presented at the annual conference of the Society for Industrial and Organizational Psychology*, Atlanta, GA.

Jagacinski, C. M., & Duda, J. L. (2001). A comparative analysis of contemporary achievement goal orientation measures. *Educational and Psychological Measurement, 61*, 1013-1039.

Kaplan, A., & Maehr, M. L. (1999). Achievement goals and student well-being. *Contemporary Educational Psychology, 24*, 330-358.

Kaplan, A., & Middleton, M. J. (2002). Should childhood be a journey or a race? A response to Harackiewicz *et al.* (2002). *Journal of Educational Psychology, 94*, 646-648.

Kaplan, A., & Midgley, C. (1999). The relationship between perceptions of the classroom goal structure and early adolescents' affect in school: The mediating role of coping strategies. *Learning and Individual Differences, 11*, 187-212.

Karabenick, S. A. (2003). Seeking help in large college classes: A person-centered approach. *Contemporary Educational Psychology, 28*, 37-58.

Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement, 3*, 103-122.

Linnenbrink, E. A. (2005). The dilemma of performance-approach goals: The use of multiple goal contexts to promote students' motivation and learning. *Journal of Educational Psychology, 97*, 197-213.

Meece, J. L. (1991). The classroom context and students' motivational goals. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 7). Greenwich, CT: JAI Press.

Meece, J. L., Blumenfeld, P., & Hoyle, R. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology, 80*, 514-523.

Meyer, D. K., Turner, J. C., & Spencer, C. A. (1997). Challenge in a mathematics classroom: Students' motivation and strategies in project-based learning. *Elementary School Journal, 97*, 501-521.

Middleton, M. J., & Midgley, C. (1997). Avoiding the demonstration of lack of ability: An unexpected aspect of goal theory. *Journal of Educational Psychology, 89*, 710-718.

Midgley, C., Kaplan, A., Middleton, M., Maehr, M. L., Urdan, T., Anderman, L. H., *et al.* (1998). The development and validation of scales assessing students' achievement goal orientations. *Contemporary Educational Psychology, 23*, 113-131.

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., *et al.* (2000). *Manual for the patterns of adaptive learning scales (PALS)*. Ann Arbor: University of Michigan.

Nadler, A. (1998). Relationship, esteem, and achievement perspectives on autonomous and dependent help seeking. In S. A. Karabenick (Ed.), *Strategic help seeking: Implications for learning and teaching* (pp. 61-93). Mahwah, NJ: Erlbaum.

Nolen, S. B., & Haladyna, T. M. (1990). Personal and environmental influences on students' beliefs about effective study strategies. *Contemporary Educational Psychology, 15*, 116-130.

Phillips, J. M., & Gully, S. M. (1997). Role of goal orientation, ability, need for achievement, and locus of control in the self-efficacy and goal-setting process. *Journal of Applied Psychology, 82*, 792-802.

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research, 31*, 459-470.

Pintrich, P. R. (2000a). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology, 25*, 92-104.

Pintrich, P. R. (2000b). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology, 92*, 544-555.

Pintrich, P. R. (2000c). The role of goal orientation in self-regulated learning. In M. Boekarts, P. R. Pintrich, & M. Zeider (Eds.), *Handbook of self-regulation* (pp. 451-502). San Diego, CA: Academic Press.

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

Roeser, R. W., Midgley, C., & Urdan, T. C. (1996). Perceptions of the school psychological environment and early adolescents' psychological and behavioral functioning in school: The mediating role of goals and belonging. *Journal of Educational Psychology, 88*, 408-422.

Ryan, A. M., & Pintrich, P. R. (1998). Achievement and social motivational influences on help seeking in the classroom. In S. A. Karabenick (Ed.), *Strategic help seeking: Implications for learning and teaching* (pp. 117-139). Mahwah, NJ: Erlbaum.

Schmidt, A. M., & Ford, J. K. (2003). Learning with a learner control-training environment: The interactive effects of goal orientation and metacognitive instruction on learning outcomes. *Personnel Psychology, 56*, 405-429.

Shih, S. (2005). Role of achievement goals in children's learning in Taiwan. *Journal of Educational Research, 98*, 310-319.

Skaalvik, E. M. (1997). Self-enhancing and self-defeating ego orientation: Relations with task and avoidance orientation, achievement, self-perceptions, and anxiety. *Journal of Educational Psychology, 89*, 71-81.

Smith, M., Duda, J. L., Allen, J., & Hall, H. (2002). Contemporary measures of approach and avoidance goal orientations: Similarities and differences. *British Journal of Educational Psychology, 72*, 155-190.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). New York, NY: Harper Collins.

Urdan, T. (2004). Using multiple methods to assess students' perceptions of classroom goal structures. *European Psychologist, 9*, 222-231.

Urdan, T., & Midgley, C. (2003). Changes in the perceived classroom goal structure and pattern of adaptive learning during early adolescence. *Contemporary Educational Psychology, 28*, 524-551.

VandeWalle, D. (1997). Development and validation of a work domain goal orientation instrument. *Educational and Psychological Measurement, 57*, 995-1015.

VandeWalle, D., Brown, S. P., Cron, W. L., & Slocum, J. W., Jr (1999). The influence of goal orientation and self-regulation tactics on sales performance: A longitudinal field test. *Journal of Applied Psychology, 84*, 249-259.

VandeWalle, D., Cron, W. L., & Slocum, J. W., Jr (2001). The role of goal orientation following performance feedback. *Journal of Applied Psychology, 86*, 629-640.

Wahlstrom, D. A. (2001, April). The relationship between goal orientation and the calibration of performance expectations to performance feedback. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

Wilson, M. (2005). *Constructing measures: An item response modelling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Winne, P. H., Muis, K. R., & Jamieson-Noel, D. J. (2006). *Relationships among achievement goal orientation, calibration bias and performance in response to successive feedback in an undergraduate course*. Unpublished manuscript.

Wolters, C., Yu, S., & Pintrich, P. R. (1996). The relation between goal orientation and students' motivational beliefs and self-regulated learning. *Learning and Individual Differences, 8*, 211-238.

Received 17 September 2007; revised version received 9 October 2008