Educational and Psychological Measurement
Volume 69 Number 2
April 2009 266-286
© 2009 SAGE Publications
10.1177/0013164408322017
http://epm.sagepub.com
hosted at
http://online.sagepub.com

A Cross-Cultural Examination of the Psychometric Properties of Responses to the Achievement Goal Questionnaire

Kou Murayama
Tokyo Institute of Technology, Japan
Mingming Zhou
John C. Nesbit
Simon Fraser University, Canada

The psychometric properties of scores from the Achievement Goal Questionnaire were examined in samples of Japanese ($N = 326$) and Canadian ($N = 307$) postsecondary students. Previous research found evidence of a four-factor structure of achievement goals in U.S. samples. Using confirmatory factor-analytic techniques, the authors found strong evidence for the four-factor structure of achievement goals in both the Canadian and Japanese populations. Subsequent multigroup structural equation modeling indicated the metric invariance of this four-factor structure across the two populations.

Keywords: *achievement goals; validity; multigroup confirmatory factor analysis*

Achievement goal theory has emerged as a major direction in motivational research (Ames, 1992; Maehr, 1989; Nicholls, 1989; Weiner, 1990). Varied theoretical perspectives have influenced the hypothesized structure of achievement goals, and the way it is measured. At an early stage of these developments, Dweck (1986; Dweck & Leggett, 1988) proposed two broad classes of goals on a single bipolar continuum with learning goals at one end and performance goals at the other. Later, achievement goals were conceptualized as two independent dimensions (Ames & Archer, 1988), wherein learning goals (or mastery goals) are motivated by learning for its own sake and performance goals are concerned with gaining public recognition

Authors' Note: This research was partly supported by grants from the Social Sciences and Humanities Research Council of Canada to John C. Nesbit (410-2007-1157). The Canadian data analyzed in this article are drawn from a larger research project, part of which is reported in Nesbit et al. (2007). We describe here only features of the data relevant to testing our theoretical goal orientation model. Further information about the larger study is available from the authors. Please address correspondence to Mingming Zhou, Faculty of Education, Simon Fraser University, Burnaby, V5A 1S6, British Columbia, Canada; e-mail: mzhou2@sfu.ca.

of one's ability, surpassing normative-based standards, or achieving success with little effort (Meece, Blumenfeld, & Hoyle, 1988).

Identifying the distinction between seeking favorable judgments of competence and avoiding negative judgments, researchers proposed a three-factor model of achievement goals that separated performance-approach from performance-avoidance goals (e.g., Elliot & Church, 1997; Elliot & Harackiewicz, 1996; Middleton & Midgley, 1997). Individuals pursuing performance-approach goals focus on doing well relative to others and, accordingly, keep performance efforts channeled toward the normative standards that eventuate in high levels of performance (Elliot & Church, 1997; Harackiewicz, Barron, Carter, Lehot, & Elliot, 1997). In contrast, performance-avoidance goals center on the possibility of failure relative to others, evoke threat appraisal and low competence expectancies, disrupt concentration, and elicit anxiety (e.g., Elliot & McGregor, 2001). Individuals with high performance-avoidance goals are not engaged in the intrinsic aspects of the task and instead focus on avoiding demonstrations of failure (Linnenbrink & Pintrich, 2000).

Elliot and McGregor (2001) extended this framework by proposing a 2×2 conceptualization of achievement goals that bifurcates both mastery and performance goals into approach and avoidance valences. Little empirical research has been done in achievement goal research about how the mastery and avoidance components integrate and function together. Theoretically, mastery-avoidant learners are concerned with loss of mastery and are not necessarily concerned with social comparison.

As the dimensionality of achievement goals has grown, so has the number of instruments that assess it. Various measures have been created and compared to research the relationships between achievement goals and other motivational variables. In studies that did not consider a four-factor goal framework, a three-factor structure appeared to be the most appropriate, regardless what instruments were used (e.g., Day, Radosevich, & Chasteen, 2003; Elliot & Church, 1997; Midgley et al., 1998, VandeWalle, 1997). However, research using the measures that were designed to demonstrate the existence of more than three latent goal constructs has lent support to the four-factor model of achievement goals. In a confirmatory factor analysis (CFA) of their Achievement Goal Questionnaire (AGQ), Elliot and McGregor (2001) uncovered a clear four-factor structure that reflected independent dimensions of approach-avoidance and mastery-performance goals. The comparison of the four-factor model with the three-factor and two-factor models led to the conclusion that the four-factor model offered the best fit to the data. Additional support for the four-factor model was found in Finney, Pieper, and Barron's (2004) study in a general academic context with a modified version of the AGQ. Evaluations of the standardized residuals indicated that the four-factor model reproduced the relationships among all items well, with none of the alternative models showing

adequate fit indices. Most recently, Attenweiler and Moore (2006) adapted several other scales to examine achievement goals in their multiwave study (Brett & VandeWalle, 1999; Button, Mathieu, & Zajac, 1996; Elliot & Church, 1997; Harackiewicz et al., 1997; Horvath, Scheu, & DeShon, 2001; VandeWalle, Cron, & Slocum, 2001). A four-factor model was found superior to both the two-factor and three-factor models. Note, however, that their four-factor model was characterized by partitioning the performance-avoidance goal into preferring tasks previously performed successfully and fear of poor performance, which departs substantially from the model identified by Elliot and McGregor (2001) in which mastery goals were divided into approach and avoidance goals.

The observed good fit of the four-factor model has significant implications for achievement goal studies, particularly for the newly identified mastery-avoidance goal. Given the somewhat counterintuitive notion of mastery-avoidance goals and the relative lack of relevant construct validation data, the future of this goal construct is unclear. To identify individuals who adopt mastery-avoidance goals and investigate their prevalence and motivational characteristics, researchers must better understand the psychometric and contextual properties of the scores from instruments that assess this construct.

Present Study

Despite support for the four-factor model of achievement goals as summarized earlier, certain issues still remain. First, as VandeWalle (1997) observed, many studies lack extensive construct validation evidence. Especially, the statistical procedures used to assess most of the measures are confined to exploratory factor analysis (EFA) and reliability analysis using Cronbach's alpha. This constitutes a shaky basis for assuming this goal framework applies broadly. Published evidence using CFA is still lacking with the exception of Button et al. (1996), Elliot and McGregor (2001), Finney et al. (2004), and a few others. Although CFA is not an inherently superior statistical technique to EFA, it does provide complementary information to that offered by EFA (Hurley et al., 1997). Specifically, CFA allows for the direct comparison of various alternative factor structures to estimate which structure provides the best description of the data, and it supports an across-group examination of the hypothesized structure.

Second, to our knowledge, there have been no studies testing an achievement goal instrument with samples from different cultures in confirmatory analyses. Attempts to demonstrate the structure and correlates of achievement goals have been undertaken using samples mainly from English-speaking countries. Despite a few studies conducted in non-English-speaking countries (e.g., Pekrun, Elliot, & Maier, 2006; Tanaka & Yamauchi, 2001; Van Yperen, 2006), methodologically rigorous studies are still lacking, which makes it difficult to compare the results

across cultures. Given the extensive research activity on achievement goals and the fact that researchers have long insisted on the importance of cross-cultural research on achievement goal constructs (Maehr & Nicholls, 1980; Urdan, 1997, 2004), the lack of cross-cultural CFA analyses is rather surprising.

It is possible that achievement goals vary in the way they manifest across cultures (Urdan, 2004). For instance, in his review on achievement goals, Murayama (2003) observed a higher correlation between performance-approach and performance-avoidance goals in Japanese samples than in American samples, indicating that Japanese people may not sharply differentiate performance-approach goals from performance-avoidance goals. Tanaka and Yamauchi (2001) found that mastery goals were positively correlated with intrinsic interest and academic achievement with their Japanese sample, but scores on both performance-approach goals and performance-avoidance goals had no statistically significant relationships with either intrinsic interest or achievement. This does not seem to accord with the results obtained with American populations (e.g., Elliot & Church, 1997), wherein mastery goals had a positive effect on intrinsic interest in learning, performance-approach goals had a positive effect on academic achievement, and performance-avoidance goals had a negative effect on both intrinsic interest and academic achievement. Also, Tanaka, Okuno, and Yamauchi (2002) observed a direct positive link between achievement motive and performance-avoidance goals, which was not found in the studies with Western samples (e.g., Elliot & Church, 1997; Elliot & McGregor, 2001).

On the other hand, there are methodologically rigorous studies showing that the structure of goals and values is largely universal across cultures (e.g., Grouzet et al., 2005; Schwartz & Bilsky, 1990). For example, Grouzet et al. (2005) showed that the 11 types of goals the authors assessed are consistently organized in a circumplex fashion across 15 cultures including Asian (China and South Korea) as well as Western (Canada and United States) countries. Although achievement goals were not investigated in these studies, the findings imply that the 2×2 structure of achievement goals, like other goals and values, could emerge across countries with different cultural backgrounds.

These developments led us to investigate (a) the extent to which the 2×2 model of achievement goals and the alternative two- and three-factor models emerge in a Japanese sample and a Canadian sample, and (b) how the interpretations of goals are different and similar across those two samples. This cross-cultural assessment of the instrument was conducted to determine the robustness and the generalizability of the four-factor model of achievement goals. Complementing earlier studies, we conducted a confirmatory, multiple-group analysis that tested aspects of similarity such as factor pattern coefficients and factor means. We chose a sample from Canada to validate previous results from the United States (e.g., Elliot & McGregor, 2001) using data from another Anglophone society with predominantly Western cultural values. We chose a sample from Japan because cross-cultural researchers have observed that Japanese samples tend to be highly representative of Asian cultural

values and strongly differentiated from Western samples (e.g., Heine, Lehman, Markus, & Kitayama, 1999). We believe these two samples constitute a solid basis for a strong test of invariance.

Method

Participants and Procedure

A total of 326 students (36.6% female) from two private (non-international and non-Christian) universities in Japan participated in early 2006. Both Japanese universities place near the middle in nationwide rankings of Japanese universities' academic performance. They are both located in urban areas of the Kanto region. Students in one university were enrolled in an introductory educational psychology class. Students in the other university were enrolled in an introductory social psychology class.

The Canadian sample consisted of 307 students (78.8% female, 80.9% Caucasians) from a public university who were enrolled in a semester-long undergraduate course in introductory educational psychology. The Canadian university is a highly ranked comprehensive university whose recently admitted students have a mean high school leaving grade of approximately 86% (Dwyer, 2007). Data were collected during the 2004/2005 academic year.

The students in the two samples completed the questionnaire during regular classroom hours. The classes in both samples were conducted in lecture format, and evaluation was based on exam performance. The achievement levels and academic environments of these samples are similar to those in the previous studies that used the AGQ (e.g., Elliot & McGregor, 2001). All participants were assured of the confidentiality of their responses.

Instrument

This study examined the psychometric properties of the scores obtained from the AGQ developed by Elliot and McGregor (2001). The instrument was administered with a 1 (*strongly disagree*) to 7 (*strongly agree*) Likert scale, with a higher score meaning a stronger orientation toward one achievement goal. Three items of the AGQ assess one of four achievement goals (mastery-approach goals, e.g., "I want to learn as much as possible from this class"; mastery-avoidance goals, e.g., "I worry that I may not learn all that I possibly could in this class"; performance-approach goals, e.g., "It is important for me to do better than other students"; and performance-avoidance goals, e.g., "My goal in this class is to avoid performing poorly").

Translation Procedure

The original English version of the AGQ was first translated into Japanese by the first author of the present study. The translation was in accordance with Brislin's (1980) recommendations for pragmatic translation of psychometric instruments. First, the inventory was translated from English into Japanese and then transcribed back to the source language as a check for consistency. Next, the original and back-translated versions were compared to ensure the accuracy of content. Finally, idiomatic changes were made to ensure comprehension for Japanese students. In the course of the process, we found that the Japanese translation of one performance-avoidance goal item ("My fear of performing poorly in this class is often what motivates me") read very unnaturally in Japanese. Accordingly, the first author paraphrased this item while maintaining the original meaning of the item.

Statistical Analyses

In addition to assessing internal consistency coefficients, CFA using Mplus 4.21 was conducted on covariance matrices to evaluate the original four-factor structure identified by Elliot and McGregor (2001) for the Japanese and Canadian populations. Given the acceptable fit of the four-factor model for each group, we used multigroup CFAs to explore whether the identified goal structure holds across two samples. Multigroup CFAs were performed by testing separate nested CFA models, including the configural invariance model, metric invariance model, factor covariance invariance model, factor variance invariance model, and scalar invariance model, with each model adding restrictions to the differences allowed across the two groups (Brown, 2006). This method has been used in many cross-cultural studies to examine construct comparability in a more rigorous way (e.g., Byrne & Campbell, 1999). The error variance invariance model was excluded in this study because it is now widely accepted that testing for the invariance of these error parameters represents an overly restrictive test of the data (Byrne, 2004). Given the fact that only 2.8% of the cases contained missing values, cases with missing values were deleted listwise in all the analyses.

Results

Descriptive Statistics and Reliability

Table 1 displays the means, standard deviations, and internal consistency coefficients for each of the analyzed variables in both samples. Acceptable thresholds for internal consistency reliability (alpha) are typically set at .70 (Nunnally, 1978) or .80 (Henson, 2001). High values for Cronbach's coefficient, ranging from .90 to .96,

Table 1
Descriptive Statistics and Reliabilities in Two Samples

Sample	Variable	*M*	*SD*	α	95% Confidence Intervals
Japan ($N = 326$)	Mastery-approach goals	14.76	3.37	.76	(0.707, .799)
	Mastery-avoidance goals	10.96	3.81	.77	(0.717, .806)
	Performance-approach goals	11.19	4.09	.79	(0.747, .827)
	Performance-avoidance goals	10.90	3.89	.72	(0.660, .768)
Canada ($N = 307$)	Mastery-approach goals	15.39	3.73	.91	(0.896, .930)
	Mastery-avoidance goals	12.50	4.42	.91	(0.897, .930)
	Performance-approach goals	12.92	4.65	.96	(0.950, .966)
	Performance-avoidance goals	13.17	4.74	.90	(0.880, .918)

Note: M = mean; SD = standard deviation. The 95% confidence intervals for Cronbach's coefficient alpha were calculated using a method employing the central F distribution (see Fan & Thompson, 2001).

were reported for scores from the Canadian group, and between .70 and .80 for the Japanese group. This indicated sufficiently high score reliability, based on classical test theory, in both samples. We then assessed skewness and kurtosis values for each item as well as multivariate normality for all the items. For both samples, item distributions are basically left-skewed and have light tails and flatness (DeCarlo, 1997) with a few exceptions (for the Canadian sample, skewness ranged from −.64 to .02 with a median of −.43; for the Japanese sample, skewness ranged from −.58 to .69 with a median of −0.11). Mardia's (1970) coefficient of multivariate normality for Canadian and Japanese samples is 27.23 and 40.17, respectively, indicating that the assumption of multivariate normality is not tenable for both samples.

Confirmatory Factor Analyses

To evaluate the structure of achievement goals across the countries, a series of CFAs were conducted (Brown, 2006). To correct for the potential statistical biases resulting from the nonnormality, maximum likelihood parameter estimates with robust standard errors and mean-adjusted chi-square test statistic (Muthén & Muthén, 2004) were used. As suggested by Hu and Bentler (1999; see also Hu & Bentler, 1998), the overall fit of the models was assessed by combinational rules of standardized root mean squared residual (SRMR) $\leq .10$ and comparative fit index (CFI) $\geq .95$ rather than less rigorous conventional criteria (e.g., CFI $> .90$), given the fact that the AGQ is a well-studied and widely used instrument. In addition to these fit indices, Tucker–Lewis index (TLI) and root mean square error of approximation (RMSEA; Browne & Cudeck, 1993) were also reported as supplements. When we compared the nested models, the difference in fit was also evaluated by CFI and TLI (G. Cheung & Rensvold, 2002; Little, 1997).

Separate CFA analyses. To address the first research question, we began by conducting CFAs separately for each country to see whether the postulated four-factor goal structure is tenable in each of the two groups. In the four-factor model, the items for each dimension loaded on their respective latent factors with no correlated errors. Prior to the analyses, a series of global tests of the equal mean and covariance structures across some demographic variables (e.g., university for the Japanese sample, ethnicity for the Canadian sample, and gender for both samples) were conducted to ensure that the mean and the covariance structures for these subgroups are homogeneous within each country (Jöreskog, 1971). For the Japanese sample, the mean and the covariance structures across genders or universities did not differ substantially: $\chi^2(90, N = 307) = 121.42$, $p < .05$, SRMR = .070, CFI = .97, TLI = .96, RMSEA = .048 for the gender comparison; $\chi^2(90, N = 308) = 145.01$, $p < .01$, SRMR = .080, CFI = .95, TLI = .93, RMSEA = .063 for the university comparison. The Canadian sample also showed relative equivalence of the mean and the covariance structures across gender or ethnicity (categorized as Caucasian or non-Caucasian): $\chi^2(90, N = 307) = 90.66$, $p = .46$, SRMR = .058, CFI = 1.00, TLI = 1.00, RMSEA = .007 for the gender comparison; $\chi^2(90, N = 307) = 139.60$, $p < .01$, SRMR = .092, CFI = .98, TLI = .98, RMSEA = .060 for the ethnicity comparison. Based on these results, we decided to pool these subgroups for each country in all the subsequent analyses.

The CFA results from the Japanese group supported the hypothesized four-factor model (see Table 2) with a good fit to the data, $\chi^2(48, N = 308) = 93.21$, $p < .01$, SRMR = .055, CFI = .96, TLI = .94, RMSEA = .055, with high factor loadings (ranging from .54 to .94). The Canadian sample also showed acceptable fit, $\chi^2(48, N = 307) = 135.72$, $p < .01$, SRMR = .048, CFI = .97, TLI = .96, RMSEA = .077, but a moderately high RMSEA value suggested that the model could be further improved. Indeed, modification indices indicated that allowing covariance between the mastery-avoidance goal factor and one error term of the performance-avoidance goal item (''My fear of performing poorly in this class is often what motivates me'') could improve the model fit substantially (modification index = 33.48; expected parameter change = 0.29). This model showed very good fit to the data (see Table 2), $\chi^2(48, N = 307) = 98.21$, $p < .01$, SRMR = .043, CFI = .98, TLI = .98, RMSEA = .060, with high factor loadings (ranging from .80 to .96) and, therefore, was determined to be used in the following analyses for the Canadian sample.

Additional CFAs were also conducted to compare the fit of all the alternative three-factor models with the hypothesized four-factor model (see Elliot & McGregor, 2001) in two samples separately. Four alternative trichotomous models were created and compared: (a) Model A, in which performance-approach and performance-avoidance items loaded on their respective latent factors, and both mastery-approach and mastery-avoidance items loaded together on a third latent factor; (b) Model B, in which mastery-approach and mastery-avoidance items loaded on their respective

Table 2
Goodness-of-Fit Values for Different Models in Two Samples

	Japanese Sample						Canadian Sample					
Models	χ^2	*df*	SRMR	CFI	TLI	RMSEA	χ^2	*df*	SRMR	CFI	TLI	RMSEA
Four-factor model	93.21	48	.055	.96	.94	.055	98.21	47	.043	.98	.98	.060
A	303.08	51	.114	.76	.68	.127	644.76	50	.160	.80	.74	.197
B	239.26	51	.106	.82	.76	.109	567.99	50	.141	.83	.77	.184
C	320.83	51	.131	.74	.66	.131	1107.87	50	.200	.65	.54	.263
D	323.91	51	.138	.73	.66	.132	516.11	50	.111	.85	.80	.174

Note: *df* = degrees of freedom; SRMR = standardized root mean squared residual; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation. Results for the Canadian sample assumed a covariance between mastery-avoidance goal factor and an error term of a performance-avoidance goal item. All *p* values for the chi-square test are less than .01.

latent factors, and both performance-approach and performance-avoidance items on a third latent factor; (c) Model C, in which mastery-approach and performance-approach items loaded on their respective latent factors, and mastery-avoidance and performance-avoidance items on a third latent factor; (d) Model D, in which mastery-avoidance and performance-avoidance items loaded on their respective latent factors, and mastery-approach and performance-approach items on a third factor. These alternative models were constructed by setting the covariance between the two combined factors to 1.0 and their covariances with other factors to be equal. For example, when performance-approach and performance-avoidance were combined (Model A), the covariance of performance-approach and mastery-approach was set to equal the covariance of performance-avoidance and mastery-approach, and the covariance of performance-approach and mastery-avoidance was set to equal the covariance of performance-avoidance and mastery-avoidance. The rationale is that because the combined factors are regarded as equivalent their covariance with other factors should be the same.

As seen in Table 2, none of the three-factor models reported a good fit to the data: $\chi^2(51, N = 308)$ varied between 239.26 and 323.91 ($ps < .01$), SRMR between .106 and .138, CFI between .73 and .82, TLI between .66 and .76, RMSEA between .109 and .132 for the Japanese population; $\chi^2(50, N = 307)$ between 516.11 and 1107.87 ($ps < .01$), SRMR between .111 and .200, CFI between .65 and .85, TLI between .54 and .80, RMSEA between .174 and .263 for the Canadian population.

Overall, these results indicated that the four-factor solution best accounted for the structure and organization of constructs measured by the AGQ with the exception of the covariance between mastery-avoidance goal factor and one error term of the performance-avoidance goal item in the Canadian sample.

Omnibus test of the mean and the covariance structure. Our second research question was to examine whether the four-factor goal structure is equivalent across the two samples. To address this issue, a series of multigroup CFAs were conducted by imposing equivalence constraints at each of several increasingly stringent levels. Prior to these analyses, however, we began with an overall test of equal means and covariance matrices across groups as a preliminary analysis (Jöreskog, 1971). Results clearly indicated that the means and covariance matrices cannot be regarded as invariant across the countries, with very bad fit to the data: $\chi^2(90, N = 615) = 895.96$, $p < .01$, SRMR = .136, CFI = .80, TLI = .70, RMSEA = .171. Therefore, we proceeded to conduct a series of increasingly restrictive multigroup CFAs to identify the source of nonequivalence. It should be noted that whether multigroup CFAs should be prefaced by the overall test has been an issue of debate, as some researchers have questioned the utility of the omnibus test (e.g., Byrne, 1998) whereas other researchers have supported its use (e.g., Vandenberg & Lance, 2000). Despite their different perspectives, however, both would agree that our results obviously suggested substantial differences in some aspects of the structure of achievement goals between the Japanese and Canadian samples.

Configural invariance model. The configural invariance model (Horn, McArdle, & Mason, 1983) is the model in which the same pattern of fixed (zero) and free factor loadings is specified for each group (Steenkamp & Baumgartner, 1998). This model is the "minimal condition for factorial invariance" (Marsh, 1993, p. 851) and provides the basis for comparison with all subsequent models (Hong, Malik, & Lee, 2003). As specified in the first CFA analysis, for the Canadian sample, the covariance between the mastery-avoidance goals and one error term of the performance-avoidance goal items remained unconstrained (i.e., freely estimated). The fit of this model was satisfactory (see Table 3), $\chi^2(95, N = 615) = 190.96$, $p < .01$, SRMR = .049, CFI = .98, TLI = .97, RMSEA = .057, and all the estimated factor loadings were quite high (ranging from .54 to .96), providing additional evidence that the scores obtained from the AGQ items exhibited the same configuration of factor pattern coefficients in Canadian and Japanese students.

Metric invariance model. Metric invariance was tested by constraining the matrix of factor pattern coefficients to be identical across groups. Again, for the Canadian sample, the covariance between the mastery-avoidance goals and one of the performance-avoidance goal items was freely estimated. As seen in Table 3, results showed a good fit to the data, $\chi^2(103, N = 615) = 263.99$, $p < .01$, SRMR = .069, CFI = .96, TLI = .95, RMSEA = .071. The differences of the fit were marginally acceptable, scale-corrected (Satorra & Bentler, 2001) $\Delta\chi^2(8, N = 615) = 60.99$, $p < .01$, ΔCFI = .02, ΔTLI = .02, suggesting that all the factor loadings could be considered as invariant across the countries.

Table 3
Model Tested Regarding the Factor Structure Across the Two Samples

Models	Model Fit								
	χ^2	*df*	SRMR	CFI	TLI	RMSEA	$\Delta\chi^2$	ΔCFI	ΔTLI
1. Configural invariance	190.96	95	.049	.98	.97	.057	—	—	—
2. Metric invariance	263.99	103	.069	.96	.95	.071	60.99	.02	.02
3a. Factor covariance invariance	305.38	109	.105	.95	.94	.077	37.58	.01	.01
3b. Factor variance invariance	298.42	107	.125	.95	.94	.076	44.56	.01	.01
3c. Scalar invariance	527.78	111	.094	.89	.87	.111	299.82	.07	.08
3d. Final metric invariance with partial scalar invariance	317.23	108	.071	.95	.94	.079	50.98	.01	.01

Note: *df* = degrees of freedom; SRMR = standardized root mean squared residual; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation. All chi-square values were scale corrected. To calculate the differences in model fit, Model 2 was compared with Model 1; Models 3a, 3b, 3c, and 3d were compared with Model 2. All *p* values for the chi-square test are less than .01.

Factor covariance and factor variance invariance models. Given that the metric invariance model was supported, we further investigated the more restricted model. First, the factor covariance invariance model was tested, in which factor covariances were set to be equal across the groups. Our results, however, did not support the model, with an SRMR being larger than .10, $\chi^2(109,\ N = 615) = 305.38,\ p < .01$, SRMR = .105, CFI = .95, TLI = .94, RMSEA = .077, although the differences of the fit between the factor covariance invariance model and the metric invariance model showed good fit, scale-corrected $\Delta\chi^2(6,\ N = 615) = 37.58,\ p < .01$, ΔCFI = .01, ΔTLI = .01. These results indicated that the relationship between achievement goals were not invariant across the countries. We also tested the factor variance invariance model, in which factor variances were constrained to be equal across the groups. The model again showed a bad fit to the data, with an SRMR being larger than .10, $\chi^2(107,\ N = 615) = 298.42,\ p < .01$, SRMR = .125, CFI = .95, TLI = .94, RMSEA = .076, although the differences of the fit between the variance invariance model and the metric invariance model showed good fit, scale-corrected $\Delta\chi^2(4,\ N = 615) = 44.56$, $p < .01$, ΔCFI = .01, ΔTLI = .01. This result suggested that the variances of the latent factors were not equivalent across the Japanese and Canadian samples.

Scalar invariance model. The preceding models require only information about the covariation of the items in different countries. To facilitate the understanding of the mean difference, the scalar invariance model was tested, in which all the item intercepts were constrained to be equal in addition to the metric invariance. Factor means for the Japanese sample were all fixed to zero to identify the model (Hancock, 1997). The overall fit for this model was not good, $\chi^2(111,\ N = 615) = 527.78$,

$p < .01$, SRMR = .094, CFI = .89, TLI = .87, RMSEA = .111. The differences of the fit between the scalar invariance model and the metric invariance model also showed a substantial deterioration, scale-corrected $\Delta\chi^2$ (8, $N = 615$) = 251.44, $p < .01$, ΔCFI = .07, ΔTLI = .08, indicating that there were substantial differences in the item intercepts as a whole across the groups.

It is often argued that proceeding with the latent means comparison is questionable and, thus, not recommended when the scalar invariance model does not indicate a good fit. Steenkamp and Baumgartner (1998), however, noted that at least one item besides the marker item has to have invariant factor loadings and invariant intercepts for cross-national comparisons of factor means to be meaningful (see also Byrne, Shavelson, & Muthén, 1989). Accordingly, based on the modification indices, we relaxed some cross-group equality constraints on the item intercepts to test the partial scalar invariance model. Specifically, modification indices suggested that substantial improvements in model fit could be obtained by relaxing the cross-group equality constraints on the intercepts of one mastery-avoidance goal item (''Sometimes I'm afraid that I may not understand the content of this class as thoroughly as I'd like''), one performance-approach goal item (''It is important for me to do better than other students''), and one performance-avoidance goal item (''I just want to avoid doing poorly in this class''). Testing of this model yielded a well-fitting model in terms of both the overall fit, $\chi^2(108, N = 615) = 317.23$, $p < .01$, SRMR = .071, CFI = .95, TLI = .94, RMSEA = .079, and the comparison of the fit between this model with that for the metric invariance model, scale-corrected $\Delta\chi^2(5, N = 615) = 50.98$, $p < .01$, ΔCFI = .01, ΔTLI = .01.

Given that more than one intercept per factor was shown to be equal across the groups, we proceeded to investigate the latent mean differences between groups (Steenkamp & Baumgartner, 1998). Statistically significant differences indicated that, for all four achievement goals, the average scores of the latent factors were higher in the Canadian sample than in the Japanese sample (mastery-approach goals, $z = 2.31$, $p < .05$; mastery-avoidance goals, $z = 2.98$, $p < .01$; performance-approach goals, $z = 7.25$, $p < .01$; performance-avoidance goals, $z = 3.23$, $p < .01$). Further examination of the mean difference effect size (Henson, 2006) revealed that this difference was especially notable in the latent means of performance-approach goals (mastery-approach goals, $d = 0.19$; mastery-avoidance goals, $d = 0.20$; performance-approach goals, $d = 0.45$; performance-avoidance goals, $d = 0.23$).

Therefore, our final model was determined to be the metric invariance model with partial scalar invariance, which is presented in Table 4 (factor patterns) and in Table 5 (factor correlations). We also calculated structure coefficients in Table 4, which represented the correlation between observed variables and latent variables (see Graham, Guthrie, & Thompson, 2003). Results showed that structure coefficients between the variables and the corresponding latent factors were always greater than the other structure coefficients, suggesting that each variable purportedly correlated with the corresponding latent factor.

Table 4
Multiple-Group Confirmatory Factor Analysis Results:
The Final Metric Invariance Model With Partial Scalar Invariance

Item	Mastery-Approach Goals		Mastery-Avoidance Goals		Performance-Approach Goals		Performance-Avoidance Goals	
	Factor Pattern	*r*	Factor Pattern	*r*	Factor Pattern	*r*	Factor Pattern	*r*
I want to learn as much as possible from this course.	1.00[a] (.69, .89)	.69, .89		.24, .44		.21, .04		−.22, −.08
It is important for me to understand the content of this course as thoroughly as possible.	1.05 (.75, .94)	.75, .94		.27, .46		.23, .04		−.24, −.09
I desire to completely master the material presented in this course.	1.11 (.70, .84)	.70, .84		.25, .41		.21, .03		−.22, −.08
I worry that I may not learn all that I possibly could in this course.		.26, .44	1.00[a] (.72, .89)	.72, .89		.20, .06		.18, .38
I am often concerned that I may not learn all that there is to learn in this course.		.25, .45	1.02 (.71, .92)	.71, .92		.20, .07		.18, .40
Sometimes I'm afraid that I may not understand the content of this course as thoroughly as I'd like.		.28, .42	1.04 (.79, .85)	.79, 85		.22, .06		.20, .37
It is important for me to do better than other students.		.20, .04		.19, .07	1.00[a] (.67, .94)	.67, .94		.35, .38

(continued)

Table 4 (continued)

Item	Mastery-Approach Goals		Mastery-Avoidance Goals		Performance-Approach Goals		Performance-Avoidance Goals	
	Factor Pattern	*r*	Factor Pattern	*r*	Factor Pattern	*r*	Factor Pattern	*r*
It is important for me to do well compared to other students in this course.		.25, .04		.23, .07	1.06 (.82, .96)	.82, .96		.43, .39
My goal in this course is to get a better score than most of the other students.		.26, .04		.24, .07	1.10 (.85, .93)	.85, .93		.45, .38
My fear of performing poorly in this class is often what motivates me.		−.21, −.08		.17, .35		.34, .33	1.00[a] (.65, .82)	.65, .82
My goal in this class is to avoid performing poorly.		−.22, −.09		.18, .40		.37, .37	1.12 (.70, .92)	.70, .92
I just want to avoid doing poorly in this class.		−.22, −.08		.18, .37		.37, .35	1.11 (.70, .87)	.70, .87

Note: *r* = structure coefficients for Japanese (left-hand side) and Canadian (right-hand side) samples. Values in the parentheses are standardized values in the Japanese (left-hand side) and Canadian samples (right-hand side), respectively.

a. Fixed values to identify the model.

Table 5
Interfactor Correlations for the Two Samples

	1.	2.	3.	4.
1. Mastery-approach goals	—	.49	.04	−.10
2. Mastery-avoidance goals	.36	—	.07	.43
3. Performance-approach goals	.30	.28	—	.41
4. Performance-avoidance goals	−.32	.26	.53	—

Note: Values above the diagonal were for the Canadian sample; values below the diagonal were for the Japanese sample.

Discussion

This study was undertaken to examine the 2×2 model of achievement goals with AGQ (Elliot & McGregor, 2001) via multigroup CFAs to investigate its invariance across two different populations. Our separate CFA findings confirmed the four-factor model of achievement goals in different populations in comparison with a series of alternative three- and two-factor models. These results indicated that the full 2×2 framework functioned well with college students of different cultural backgrounds, an important consideration in evaluating the potential utility of the theory in cultural groups other than those in which they were developed. Furthermore, subsequent multigroup CFAs revealed the differences as well as similarities of these factor structures in more detail.

The similarities of the goal structures between groups were evidenced by the metric invariance between the Japanese and Canadian samples. This implies that common factors could, though not conclusively, have the same meaning among these groups as reflected in the invariant factor loadings (direct relationships between common factors and observed item scores). Theoretically, a metric invariance model makes it possible to meaningfully compare item difference scores across countries, and these observed item differences are indicative of similar cross-national differences in the underlying construct (Steenkamp & Baumgartner, 1998). The issue of measurement invariance is profound because the comparison of scores between groups is meaningless when those scores represent qualitatively different constructs (Hoyle, 2005). As Horn (1991) argued, "Without evidence of measurement invariance, the conclusions of a study must be weak" (p. 119).

These findings are especially interesting given the fact that, since Maehr and Nicholls's (1980) early work, some researchers have argued that achievement goals might operate differently for members of collectivist (e.g., Asian) cultures than for members of individualistic (e.g., North American) cultures (Urdan, 2004; Urdan & Mestas, 2006). Urdan (2004) further explained that this cultural difference should be salient in both forms of performance goals because performance goals involve a

self-conscious, ego-involved concern with how one appears relative to others, and according to Markus and Kitayama (1991), the notion of "self" or "ego" is defined differently in collectivistic and individualistic cultures. Individualistic students are believed to be motivated primarily by the goal of feeling personal pride, whereas the collectivist is believed to be motivated by fear of feeling shame (Markus & Kitayama, 1991). Elliot, Chirkov, Kim, and Sheldon's (2001) study complemented this line of research by demonstrating that collectivists, relative to individualists, engage in more avoidance regulation at the goal level of analysis.

Our results, on the contrary, suggest that the items could be interpreted similarly between Japanese and Canadian samples at least in the sense that individual differences in the underlying constructs explain the observed item scores in the same way in the two cultures. This is in accordance with the findings of the universality of goals and values (Grouzet et al., 2005; Schwartz & Blisky, 1990) and might pose some questions to the above-described cultural relativist view of achievement goals.

Supporting the metric invariance model, however, does not mean that every psychometric aspect of the scores from AGQ is equivalent between the groups. First, our analyses did not support factor-variance and factor-covariance invariance models, although they achieved a modestly good fit except for SRMR, indicating that interfactor correlations were not the same between the two samples. Of particular interest here is the difference in the correlation between performance-approach and performance-avoidance goals. As described above, Murayama (2003) observed a higher correlation between them in Japanese samples than in American samples. Our result is consistent with his findings because it shows a statistically significant difference between correlations in the two samples, with a higher correlation between the two goals in the Japanese sample than in the Canadian sample (for Japanese, $r = .53$; for Canadians, $r = .41$; $z = 1.97$, $p < .05$). Notably, our results were obtained for the latent factor correlation in the metric invariance model, indicating that the difference cannot be attributed to the differences in the reliability of the scores or the factor loadings. Future research is needed to account for this finding.

Second, for all four achievement goals, the average scores of the latent factors were higher in Canadian students than in Japanese students. This might reflect the tendency of Canadian students to endorse all types of achievement goals in comparison with Japanese students. Recall that the questionnaire was implemented as part of the course requirement for Canadian students. We speculate that social desirability bias may have led the Canadian respondents to answer so as to convey an impression of an active and engaged motivational stance. This may explain why the average score of performance-approach goals, in particular, was higher in the Canadian sample. Also consistent with this explanation are the studies showing that Japanese exhibit less self-enhancement than North Americans (for a review, see Heine et al., 1999). In sum, it appears that systematic response bias is a strong confounding variable that makes it difficult to compare the mean differences between

cultural groups (M. Cheung & Chan, 2002; Peng, Nisbett, & Wong, 1997; see also Heine, Lehman, Peng, & Greenholtz, 2002).

Third, an observable difference in the internal consistency coefficients of reliability (Cronbach's alpha ranges from .72 to .79 for Japanese and from .90 to .96 for Canadians) indicates that the amount of measurement error was not identical in the Canadian and Japanese populations and that the items were not equally reliable across countries. This is also apparent from the standardized factor loadings (Table 4). This result demonstrates the practical importance of controlling measurement error (e.g., using the latent variable approach) when comparing the relationship of achievement goal scores obtained from AGQ with criterion variables across cultures. Finally, an unexpected positive correlation between mastery-avoidance goal factor and an error term of one performance-avoidance goal item ("My fear of performing poorly in this class is often what motivates me") was observed in the Canadian sample. This might be produced by the affective components (e.g., fear or anxiety) contained in both constructs, and ideally, these affective components should be excluded from the items in the future revision of the scale as was done by Cury, Elliot, Da Fonseca, and Moller (2006) or Elliot and Murayama (in press).

It is essential for scientific inference to have evidence of measurement equivalence (Steenkamp & Baumgartner, 1998), especially in cross-cultural studies (Matsumoto & Yoo, 2006). Such evidence, unfortunately, is often not presented in cross-national research in the behavioral sciences (Horn & McArdle, 1992; Hui & Triandis, 1985). The lack of such evidence has led to equivocal conclusions, casting doubt on their implications for theory (Horn & McArdle, 1992). In this study, we were able to demonstrate the equivalence of the four-factor structure of AGQ within a diversified sample of respondents and show that the psychometric properties of its scores can be similar to some extent across culturally and linguistically different populations. But our examination of the cross-cultural factor structures of AGQ is only one key step toward cross-cultural validation of the measurement. To convincingly validate a measurement that supposedly reflects hypothetical constructs such as achievement goals, the measure of a given construct has to relate to measures of other constructs in theoretically predictable ways (Cronbach & Meehl, 1955). That is, correlations with theoretically relevant criteria also constitute crucial evidence of validity (Clark & Watson, 1995). Thus, we recommend that construct validation methods assessing criterion validity across cultures should be used in future work to determine the functioning of the 2×2 goal model in different cultures.

References

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*, 260-267.

Attenweiler, W. J., & Moore, D. (2006). Goal orientations: Two, three, or more factors? *Educational and Psychological Measurement*, *66*, 342-352.

Brett, J. F., & VandeWalle, D. (1999). Goal orientation and goal content as predictors of performance in a training program. *Journal of Applied Psychology*, *84*, 863-873.

Brislin, R. W. (1980). Translation and content analysis of oral and written materials. In H. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology* (Vol. 2., pp. 389-444). Boston: Allyn & Bacon.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Beverley Hills, CA: Sage.

Button, S. B., Mathieu, J. E., & Zajac, D. M. (1996). Goal orientation in organizational research: A conceptual and empirical foundation. *Organizational Behavior and Human Decision Processes*, *67*, 26-48.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS and SIMPLIS: Basic concepts, applications and programming*. Mahwah, NJ: Lawrence Erlbaum.

Byrne, B. M. (2004). Testing for multigroup invariance using AMOS graphics: A road less traveled. *Structural Equation Modeling*, *11*, 272-300.

Byrne, B. M., & Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-Cultural Psychology*, *30*, 555-574.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255.

Cheung, M. W. L., & Chan, W. (2002). Reducing uniform response bias with ipsative measurement in multiple group confirmatory factor analysis. *Structural Equation Modeling*, *9*, 55-77.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309-319.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

Cury, F., Elliot, A. J., Da Fonseca, D., & Moller, A. (2006). The social-cognitive model of achievement motivation and the 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, *90*, 666-679.

Day, E. A., Radosevich, D. J., & Chasteen, C. S. (2003). Construct- and criterion-related validity of four commonly used goal orientation instruments. *Contemporary Educational Psychology*, *28*, 434-464.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, *2*, 292-307.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*, 1040-1048.

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

Dwyer, M. (2007). Our 17th annual rankings. *Maclean's, 120*(45), 100-116.

Elliot, A. J., Chirkov, V. I., Kim, Y., & Sheldon, K. M. (2001). A cross-cultural analysis of avoidance (relative to approach) personal goals. *Psychological Science*, *12*, 505-510.

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

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

Elliot, A. J., & Murayama, K. (in press). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology*.

Fan, X., & Thompson, B. (2001). Confidence intervals about score reliability coefficients, please: An EPM guidelines editorial. *Educational and Psychological Measurement*, *61*, 517-531.

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

Graham, J. M., Guthrie, A. C., & Thompson, B. (2003). Consequences of not interpreting structure coefficients in published CFA research: A reminder. *Structural Equation Modeling*, *10*, 142-153.

Grouzet, F. M., Kasser, T., Ahuvia, A., Dols, J. M., Kim, Y. Lau, S., et al. (2005). The structure of goals across 15 cultures. *Journal of Personality and Social Psychology*, *89*, 800-816.

Hancock, G. R. (1997). Structural equation modeling methods of hypothesis testing of latent variable means. *Measurement and Evaluation in Counseling and Development*, *30*, 91-105.

Harackiewicz, J. M., Barron, K. E., Carter, S. M., Lehot, A. T., & Elliot, A. J. (1997). Predictors and consequences of achievement goals in the college classroom: Maintaining interest and making the grade. *Journal of Personality and Social Psychology*, *73*, 1284-1295.

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review*, *106*, 766-794.

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales: The reference-group problem. *Journal of Personality and Social Psychology*, *82*, 903-918.

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, *34*, 177-189.

Henson, R. K. (2006). Effect-size measures and meta-analytic thinking in counseling psychology research. *Counseling Psychologist*, *34*, 601-629.

Hong, S., Malik, M. L., & Lee, M. (2003). Testing configural, metric, scalar, and latent mean invariance across genders in sociotropy and autonomy using a non-Western sample. *Educational and Psychological Measurement*, *63*, 636-654.

Horn, J. L. (1991). Comments on "Issues in factorial invariance". In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 114-125). Washington, DC: American Psychological Association.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*, 117-144.

Horn, J. L., McArdle, J., & Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *Southern Psychologist*, *1*, 179-188.

Horvath, M., Scheu, C. R., & DeShon, R. P. (2001, April). *Goal orientation: Integrating theory and measurement*. Paper presented at the annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Hoyle, R. H. (2005). Applications of structural equation modeling in clinical and health psychology research. In J. Miles & P. Gilbert (Eds.), *A handbook of research methods for clinical and health psychology* (pp. 261-278). New York: Oxford University Press.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424-453.

Hu, L., & Bentler, P. M. (1999). Cutoff criterion for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Hui, C. H., & Triandis, H. C. (1985). Measurement in cross-cultural psychology. *Journal of Cross-Cultural Psychology*, *16*, 131-152.

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., et al. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, alternatives. *Journal of Organizational Behavior*, *18*, 667-683.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409-426.

Linnenbrink, E. A., & Pintrich, P. R. (2000). Multiple pathways to learning and achievement: The role of goal orientation in fostering adaptive motivation, affect, and cognition. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 195-227). San Diego, CA: Academic Press.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*, 53-76.

Maehr, M. L. (1989). Thoughts about motivation. In C. Ames & R. Ames (Eds.), *Research on motivation in education* (Vol. 3). New York: Academic Press.

Maehr, M. L., & Nicholls, J. G. (1980). Culture and achievement motivation: A second look. In N. Warren (Ed.), *Studies in cross-cultural psychology* (Vol. 2, pp. 221-267). New York: Academic Press.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519-530.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*, 224-253.

Marsh, H. W. (1993). The multidimensional structure of academic self-concept: Invariance over gender and age. *American Educational Research Journal*, *30*, 841-860.

Matsumoto, D., & Yoo, S. H. (2006). Toward a new generation of cross-cultural research. *Perspectives on Psychological Science*, *1*, 234-250.

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

Middleton, M., & Midgley, C. (1997). Avoiding the demonstration of lack of ability: An underexplored 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.

Murayama, K. (2003). Testing of undifferentiated performance-goal hypothesis. *Developmental Studies in Social Motivation*, *2*, 3-11.

Muthén, L. K., & Muthén, B. O. (2004). *Mplus statistical analysis with latent variables: User's guide*. Los Angeles: Statmodel.

Nesbit, J. C., Winne, P. H., Jamieson-Noel, D., Code, J., Zhou, M., MacAllister, K., et al. (2007). Using cognitive tools in gStudy to investigate how study activities covary with achievement goals. *Journal of Educational Computing Research*, *35*, 339-358.

Nicholls, J. G. (1989). *The competitive ethos and democratic education*. Cambridge, MA: Harvard University Press.

Nunnally, J. C. (1978). *Introduction to psychological measurement*. New York: McGraw-Hill.

Pekrun, R., Elliot, A. J., & Maier, M. A. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology*, *98*, 583-597.

Peng, K., Nisbett, R., & Wong, N. (1997). Validity problems comparing value across cultures and possible solutions. *Psychological Methods*, *2*, 329-344.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507-514.

Schwartz, S., & Blisky, W. (1990). Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of Personality and Social Psychology*, *58*, 878-891.

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*, 78-90.

Tanaka, A., Okuno, T., & Yamauchi, H. (2002). Achievement motives, cognitive and social competence, and achievement goals in the classroom. *Perceptual and Motor Skills*, *95*, 445-458.

Tanaka, A., & Yamauchi, H. (2001). A model for achievement motives, goal orientations, intrinsic interest, and academic achievement. *Psychological Reports*, *88*, 123-135.

Urdan, T. (1997). Achievement goal theory: Past results, future directions. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 99-141). Greenwich, CT: JAI Press.

Urdan, T. (2004). Predictors of academic self-handicapping and achievement: Examining achievement goals, classroom goal structures, and culture. *Journal of Educational Psychology*, *96*, 251-264.

Urdan, T., & Mestas, M. (2006). The goals behind performance goals. *Journal of Educational Psychology*, *98*, 354-365.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4-70.

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

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

Van Yperen, N. W. (2006). A novel approach to assessing achievement goals in the context of the 2×2 framework: Identifying distinct profiles of individuals with different dominant achievement goals. *Personality and Social Psychology Bulletin*, *32*, 1432-1445.

Weiner, B. (1990). History of motivational research in education. *Journal of Educational Psychology*, *82*, 616-622.