Longitudinal Cross-Gender Factorial Invariance of the Academic Motivation Scale

Frederick M. E. Grouzet, Nancy Otis, and Luc G. Pelletier
University of Ottawa

This study examined the measurement and latent construct invariance of the Academic Motivation Scale (Vallerand, Blais, Brière, & Pelletier, 1989; Vallerand et al., 1992, 1993) across both gender and time. An integrative analytical strategy was used to assess in one set of nested models both longitudinal and cross-gender invariance, and "interactional" and "simple" tests of invariance (e.g., cross-gender invariance across waves) were conducted. Results from a sample of 643 students (322 boys and 321 girls) across a 3-year time span showed good longitudinal cross-gender factorial invariance with different degrees of invariance depending on the sample, the educational level, or both. Implications for the validity of the Academic Motivation Scale, as well as for research on factorial invariance, are discussed.

Over the past 30 years, research on student motivation has flourished and remains a contemporary and important topic in education and psychology. In education, motivation is primarily studied in relation to school performance, as well as to the choices students make about which academic activities to engage in, their persistence in continuing these activities, and the degree of effort they expend. Exploring the why of students' behavior, several motivation theorists have made a distinction between intrinsic and extrinsic motivations (e.g., Deci, Vallerand, Pelletier, & Ryan, 1991; Harter, 1981). *Intrinsic motivation* generally refers to performing an activity as an end in and of itself—for the pleasure and satisfaction derived from this participation (e.g., Deci, 1975). On the other hand, *extrinsic motivation* refers to engaging in an activity as a means to an end; that is, to obtain something positive or to avoid something negative external to the activity (Deci, 1975). However,

---

Correspondence should be addressed to Frederick M.E. Grouzet, Educational & Counseling Psychology, McGill University, 3700 McTavish, Montreal, Quebec H3A 1Y2, Canada. E-mail: frederick.grouzet@mcgill.ca

some researchers have proposed a multidimensional view of extrinsic motivation. For instance, according to self-determination theory (Deci & Ryan, 1985, 2000), there are four types of extrinsic motivation: External and introjected regulations refer to conducting an activity due to an external and internal (respectively) sense of obligation to obtain something positive or to avoid something negative beyond the activity itself (e.g., Deci & Ryan, 1985), whereas identified and integrated regulations refer to engaging in an activity by personal choice (more or less integrated in the self; see e.g., Deci & Ryan, 1985). These four types of regulatory processes fall along a self-determination continuum from the least to most self-determined motivation, that is external, introjected, identified, and integrated regulation. Intrinsic motivation is defined as the prototype of self-determined activity and thus is placed at the self-determined pole of this continuum, and amotivation (i.e., the relative absence of motivation) is placed at the opposite pole (for a complete review of self-determination theory see Deci & Ryan, 2000).

To study intrinsic and extrinsic motivation according to self-determination theory's multidimensional perspective, a great deal of effort has been made to create and validate multiconstruct scales. Along these lines, Vallerand and his colleagues have developed the Academic Motivation Scale (AMS; Vallerand et al., 1992, 1993; its French version is the "Échelle de Motivation en Éducation" or ÉMÉ; Vallerand, Blais, Brière, & Pelletier, 1989). The AMS is one of the most frequently used scales to measure intrinsic and extrinsic academic motivation. The AMS has been integrated in empirical models which incorporate both determinants (e.g., school environment, parental engagement, professors' behaviors) and consequences (e.g., dropout, school performance, interest) of academic motivation (see e.g., Guay & Vallerand, 1997; Pelletier, Séguin-Lévesque, & Legault, 2002; Vallerand, Fortier, & Guay, 1997), providing support for its construct and predictive validity. The AMS has also been validated with various populations, with males and females, with English- and French-speaking students, across various geographic areas (e.g., Canada, United States, France), and from senior high school to university level. However, the AMS has never been tested for invariance across different groups or across time. Considering that a few studies have reported differences in motivation between females and males, the examination of cross-gender invariance would allow the interpretation of whether observed differences are at the construct level (i.e., true motivational differences) or at the measurement level (i.e., differences in scaling). Furthermore, support for the AMS's longitudinal invariance would enable researchers to further test hypotheses about the development (i.e., change) in students' motivation across education levels.

Confirmatory factor analysis (CFA) framework is ideally suited for testing whether scores measure "the correct something" (rather than "nothing"; Thompson, 2003) and whether the meaning of this "something" is equivalent across samples or times. Likewise, a recent review of literature by Vandenberg and Lance (2000) indicated that interest in examining factorial invariance (FI) from a CFA

perspective is pivotal and extensive. In fact, the authors detected more than 14 different recommended procedures for testing FI. This lack of consensus indicates that FI procedures are still in development. Furthermore, Vandenberg and Lance found 67 other articles in which other researchers used various approaches to test FI, mostly across samples (i.e., gender, cultures, etc.) but also across time. However, no study has examined FI across both samples and time in only one group of nested models. Therefore, testing FI of the AMS across gender and time provides an interesting challenge for FI research. Moreover, this challenge is ever greater due to the multidimensional nature of the AMS.

The purpose of the study was to examine further aspects of validity of the AMS with elementary and junior high-school students (i.e., from 8th to 10th grade) using an integrated structural equation modeling strategy. In particular, the study simultaneously examined both longitudinal and cross-gender FI of the AMS with an analytical strategy, using the mean and covariance structures (MACS) technique (e.g., T.D. Little, 1997), and other new approaches proposed by structural equation modeling theorists working on FI (see Vandenberg & Lance, 2000). Finally, the study hoped to show the generalizability of this analytical strategy for testing both interindividual (i.e., across samples) and intraindividual (e.g., across times) FI combined.

METHODS

Participants and Procedure

The data analyzed in this study came from a large longitudinal project conducted in an Ontario Catholic French-speaking school board (Pelletier, 2004). Over a period of 3 years (from 2001 to 2003), 322 boys and 321 girls in 8th, 9th, and 10th grades completed the French version of the AMS (Vallerand et al., 1989). Mean ages for 8th, 9th, and 10th grades were 13-, 14-, and 15-years-old, respectively. The questionnaire was answered during regular class hours. For the 8th graders, questions were read aloud to ascertain a good understanding of the questions. Ninth and 10th graders answered the questions independently.

Instrument

The AMS included five subscales assessing intrinsic motivation (IM), identified regulation (ID), introjected regulation (IJ), external regulation (ER), and amotivation (AM) with 4 items each, totaling 20 items. Students were asked "Why are you going to school?" and the response choices for each item were rated on a 5-point Likert-type scale from 0 (*does not correspond at all*) to 4 (*corresponds exactly*). The items can be found in Vallerand et al. (1989, 1992).

Missing Data

In most longitudinal studies, missing data and attrition are frequent. This study was no exception. More than 30 different missing patterns were detected in each boy and girl sample. Among them, two relevant attrition patterns were found in 9th-grade and 10th-grade waves. More specifically, 80 girls and 76 boys were absent from the 9th-grade wave, and 62 girls and 72 boys were absent from the 10th-grade wave. Although the presence of these two missing patterns suggests using multisample techniques for handling missing data (e.g., Muthén, Kaplan, & Hollis, 1987), the maximum likelihood estimate of missing data was used because of the presence of other patterns and also because a multisample approach (in addition to the gender subdivision) may require a more complex design. Results from Kim and Bentler's (2002) test of homogeneity of means and covariances revealed that the missing completely at random assumption for both means and covariances was not rejected for the boy sample but was rejected for the girl sample. For the purpose of this study, it was assumed that the data was missing at random (R.J.A. Little & Rubin, 1987). Therefore, the study used Jamshidian and Bentler's (1999) EM-type (Expected Maximization) missing data procedures for multiple samples to replace them. Then a complete set of data was worked on.

Analyzing Longitudinal Cross-Sample Invariance

The following sections describe the analytical strategy that was used in the study. To provide useful information for interested researchers each aspect of the analyses is detailed. Although the analyses were applied to the study data, further applications to other data are easily possible.

Parameters in MACS framework. The MACS approach, which was proposed by T.D. Little (1997), distinguishes measurement and latent construct parameters. Measurement parameters correspond to pattern loadings (λ), intercepts (τ), and measurement errors (θ_ε). In this study, there were five subscales (i.e., intrinsic motivation, identified regulation, introjected regulation, external regulation, and amotivation) with 4 items each, totaling 20 items. Girls and boys responded to these 20 items each year for 3 years. Thus, this model involved 20(items) × 2(girls vs. boys) × 3(waves) = 120 loadings, as well as 120 intercepts, and 120 errors.

Latent construct parameters. These are standard deviations (σ), correlations (ψ),[1] and means (α). In this study, there were 5(motivational constructs) × 2(girls vs. boys) × 3(waves) = 30 latent standard deviations, as well as 30 latent

---

[1]Latent standard deviation is estimated via a regression path between the latent construct (for which variance is fixed to zero) and a phantom variable (for which variance is fixed to 1). Latent correlation is thus estimated by a covariance parameter between two phantom variables. For more information, the reader is referred to Little (1997); see also McArdle & Cattell (1994).

means. There were three types of correlations. First, there were the interrelations among the five latent motivational constructs (i.e., intrinsic motivation, identified regulation, introjected regulation, external regulation, and amotivation). The total was 10(interrelations among motivational constructs) × 2(girls vs. boys) × 3(waves) = 60 correlations. Second, for each motivational construct one can assess longitudinal stability by estimating correlations among latent factors across waves. Three waves means there were three separate stability coefficients (i.e., from 8th to 9th grade, from 9th to 10th grade, and from 8th to 10th grade), totaling 3 × 5(motivational constructs) × 2(girls vs. boys) = 30 correlations. A third type of correlation involved relations among different motivational constructs across waves (e.g., intrinsic motivation in 8th grade with external regulation in 9th grade). Because this was not the scope of this study, theoretical meaning was not associated with this last group of correlations.

Finally, in both girl and boy groups, adjacent cross-time correlations among measurement errors were estimated, totaling 20(items) × 2(adjacent pairs of errors; i.e., from 8th to 9th grade and from 9th to 10th grade) = 40 error correlations.

Standardization and identification.   When estimating a one-sample CFA model, there are two ways to standardize a latent construct (i.e., reach scale identification). A direct way is to fix the variance (i.e., $\eta = 1$, or the standard deviation [$\sigma = 1$] estimated through the path regression—see footnote 1) and the latent mean (i.e., $\alpha = 0$) of the latent constructs. An indirect way is to fix the pattern loading (i.e., $\lambda_y = 1$) and the intercept (i.e., $\tau_y = 0$) of one selected item. However, when estimating a multisample CFA model, problems may arise when the entity (especially η and λ_y,) used for standardization (i.e., fixed to unity across samples) is not actually invariant across samples (see e.g., Cheung & Rensvold, 1999; Reise, Widaman, & Pugh, 1993). Therefore, various authors have proposed various standardization procedures, which are more or less equivalent for a "global test" (i.e., full invariance test) but may be inadequate for a partial invariance test (see e.g., Cheung & Rensvold, 1999; Reise et al., 1993).

In this study, a method of standardization adapted from Reise et al.'s (1993) recommendation (see also Cheung & Rensvold, 1999; Steenkamp & Baumgartner, 1998) was used. First, the study fixed the five latent standard deviations (i.e., path regressions), as well as the five latent means of the five motivational constructs of a referent "group" (i.e., girls in eighth grade: IM_{8th-G}, ID_{8th-G}, IJ_{8th-G}, EX_{8th-G}, and AM_{8th-G}).[2] Second, the study constrained the pattern loadings (λ_1) and the inter-

---

[2]MACS (and multigroup CFA) must select one sample as a referent. In this study, girls were arbitrarily selected. In a longitudinal model, the choice of a reference point is less arbitrary and corresponds more often to the first wave (i.e., 8th grade). Therefore, the five motivational constructs associated with the 8th-grade-girl "group" were the reference points for the corresponding motivational constructs associated with the 9th-grade-girl, 10th-grade-girl, 8th-grade-boy, 9th-grade-boy, and 10th-grade-boy groups.

cepts (τ_1) of the first item of each motivational construct to be equal across samples and waves. All other measurement and latent construct parameters were freely estimated. This model was the basic model.

Testing invariance with nested sequences of models.   The FI test was then conducted as a nested-model comparison between the basic model and one in which specific parameters were constrained to be equal across samples, waves, or both. According to T.D. Little (1997), measurement and latent construct invariance were distinguished. In general, researchers distinguish four levels of measurement invariance (MI) corresponding to four nested models, as follows (from less to more constrained): (a) Configural invariance (Thurstone, 1947) is tested through the basic model (or M_{basic}) that estimates measurement and latent construct parameters without any constraints (excepted those specified previously); (b) Metric invariance (Thurstone, 1947; Meredith's, 1993, weak invariance) is tested through models in which pattern factor loadings are constrained to be equal across samples and waves—these metric models (M_{metric}) are nested in the M_{basic} models; (c) Scalar invariance (i.e., Meredith's, 1993, strong invariance) is tested through models in which pattern factor loadings and intercepts are constrained to be equal across samples and waves—these scalar models (M_{scalar}) are nested in the M_{metric} models; (d) Uniqueness invariance (i.e., Meredith's, 1993, strict invariance) is tested through models where pattern factor loadings, intercepts, and items' unique variances are constrained to be equal across samples and waves. These uniqueness models ($M_{uniqueness}$) are nested in the M_{scalar} models.

Similarly, tests of latent construct invariance (or difference/change) also involve nested models. For the three types of latent construct parameters (i.e., standard deviation, correlation, and mean) there are three corresponding models that are nested in specific measurement models or in each other: (a) Latent standard deviation invariance is tested through models nested in the M_{metric} models, and thus pattern factor loadings and latent standard deviations are constrained to be equal across samples and waves; (b) Latent correlation invariance is tested through models that are nested in the M_{metric} models, and thus pattern factor loadings and correlation are constrained to be equal across samples and waves; (c) Latent mean invariance is tested through models nested in the M_{scalar} models, and pattern factor loadings, intercepts, and latent mean are constrained to be equal across samples and waves.

A 2(samples) × 3(waves) MACS design.   Most of the cross-sample and longitudinal FI hypotheses that have been tested in the literature were tested separately. In the study, both cross-gender and longitudinal FI in one set of MACS nested models were simultaneously tested. Testing FI through a 2 × 3 design implies numerous nested model in which equality constraints are imposed on measurement and latent construct parameters, across gender, across waves, or both.

Figure 1 illustrates a 2×3 design for FI (see details later). The first part (Figure 1a) represents FI pertaining covariance structure, whereas the second part (Figure 1b) includes invariance levels pertaining mean structure.

The 2×3 design for MI is illustrated in the Figure 1b. Each column (from *A* to *C*) represents a cross-gender-invariance level, from configural to metric to scalar invariance, and corresponds to a model in which a set of cross-gender equality constraints is imposed. Each row (from 1 to 3) represents a longitudinal-invariance level, from configural to metric to scalar invariance, and corresponds to a model in which a set of equality constraints is imposed across waves. Degree-of-freedom differences (Δdf) are shown on the arrows from one model to another nested model, which correspond to the number of constraints imposed on the nested model. For example, Model 2A was built to test longitudinal metric invariance and therefore the 30 loadings (i.e., 3 items $\times$ 5 motivational constructs $\times$ 2 samples) for the 9th-grade wave, as well as the 30 loadings in the 10th-grade wave, were constrained to be equal to those in the 8th-grade wave, corresponding to a gain (or difference) of 60 *df*s in comparison with the baseline model (Model 1A). However, because of the "interactional" nature of the present invariance test, there were overlaps in constraints and degrees of freedom were therefore adjusted. For instance, in Model 2B the same 60 longitudinal metric invariance constraints as mentioned previously were imposed in association with 45 cross-gender metric invariance constraints (i.e., 3 items $\times$ 5 motivational constructs $\times$ 3 waves). Because there were 30 overlapped constraints, the difference or gain of degrees of freedom (in comparison with Model 1A) was 75 rather than $60 + 45 = 105$. The adjustment of the degrees of freedom can also be seen when one compares degree of freedom gains from Model 1B to Model 2B ($\Delta df = 30$) and from Model 1A to Model 2A ($\Delta df = 60$) or when one compares degree of freedom gains from Model 2A to Model 2B ($\Delta df = 15$) and from Model 1A to Model 1B ($\Delta df = 45$). Finally, it is worth noting that Δdf between two nested models, which are not juxtaposed in the figure (e.g., Model 2B and Model 3C), can be calculated by adding Δdfs between intermediate nested models (e.g., $45 + 30 = 75$).

This 2×3 design for MI is particularly useful for testing "simple effect" hypotheses of MI. For instance, it becomes possible to test longitudinal invariance for girls and for boys, separately (i.e., longitudinal invariance across gender). Furthermore, this approach allows for the testing of cross-gender invariance at three periods of life (i.e., 8th, 9th, and 10th grade) separately (i.e., gender invariance across waves). As well, with this integrated approach one can test new "interactional" hypotheses of MI. For instance, it becomes possible to examine the impact of cross-gender (non-) invariance on a test of longitudinal invariance and vice versa.

The 2×3 designs can be extended to include latent constructs, such as standard deviation and correlation invariance (Figure 1a; column D and E; row 4 and 5) and mean invariance (Figure 1b; column F; row 6), respectively. Similarly, for MI, the 2×3 designs for latent constructs may be useful for testing specific hypotheses.

Gender Invariance Levels

	Configural	Metric	Stand. Dev.	Correlation

Longitudinal Invariance Levels

Configural — Model 1A Baseline model — Δdf=45 → Model 1B — Δdf=15 → Model 1D — Δdf=30 → Model 1E

Δdf=60 ↓ — Δdf=30 ↓ — Δdf=30 ↓ — Δdf=30 ↓

Metric — Model 2A — Δdf=15 → Model 2B — Δdf=15 → Model 2D — Δdf=30 → Model 2E

Δdf=20 ↓ — Δdf=20 ↓ — Δdf=10 ↓ — Δdf=10 ↓

Stand. Dev. — Model 4A — Δdf=15 → Model 4B — Δdf=5 → Model 4D — Δdf=30 → Model 4E

Δdf=40 ↓ — Δdf=40 ↓ — Δdf=40 ↓ — Δdf=20 ↓

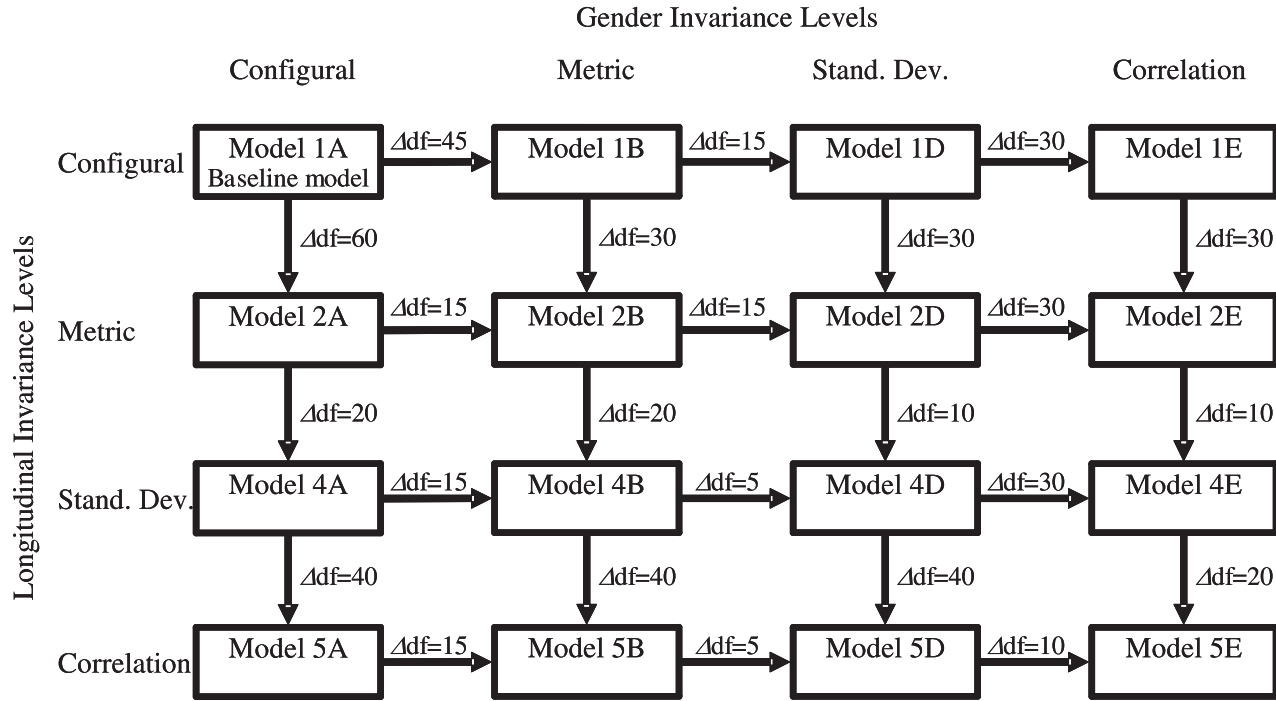Correlation — Model 5A — Δdf=15 → Model 5B — Δdf=5 → Model 5D — Δdf=10 → Model 5E

FIGURE 1   A set of nested models for testing the longitudinal cross-gender measurement invariance of the Academic Motivation Scale, the invariance of standard deviation and correlation among motivational constructs (1a) and invariance of the latent means (1b).

Gender Invariance Levels

| Configural | Metric | Scalar | Latent Mean |

Longitudinal Invariance Levels

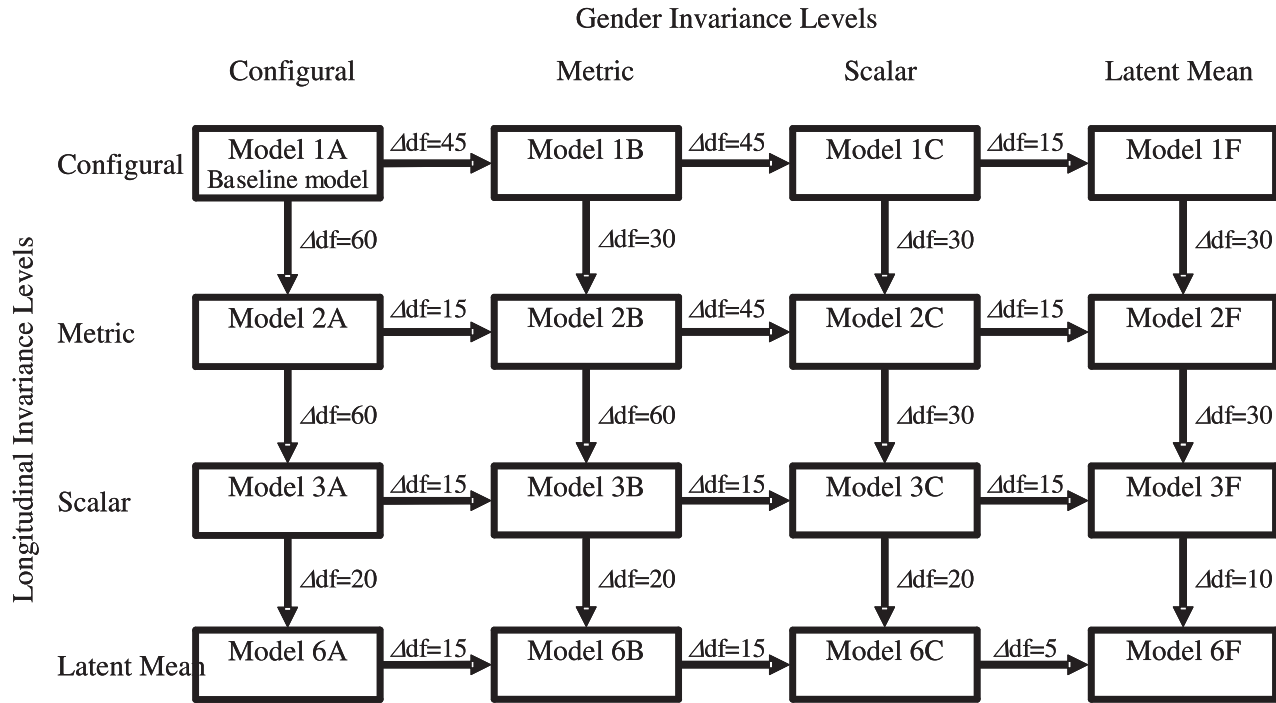|	Configural	Metric	Scalar	Latent Mean
Configural	Model 1A Baseline model	Model 1B	Model 1C	Model 1F
	Δdf=45	Δdf=45	Δdf=15	
	Δdf=60	Δdf=30	Δdf=30	Δdf=30
Metric	Model 2A	Model 2B	Model 2C	Model 2F
	Δdf=15	Δdf=45	Δdf=15	
	Δdf=60	Δdf=60	Δdf=30	Δdf=30
Scalar	Model 3A	Model 3B	Model 3C	Model 3F
	Δdf=15	Δdf=15	Δdf=15	
	Δdf=20	Δdf=20	Δdf=20	Δdf=10
Latent Mean	Model 6A	Model 6B	Model 6C	Model 6F
	Δdf=15	Δdf=15	Δdf=5	

FIGURE 1   *(Continued)*

Estimation procedure and fit indexes.   Maximum likelihood estimation procedures were used with Satorra and Bentler's (1988, 1994) scaling corrections, allowing for the calculation of the Satorra-Bentler (SB) scaled chi-square value ($SB\chi^2$).[3] Two types of fit indexes were used: overall and comparative fit indexes (see Bollen, 1989). The overall fit indexes used in the study were the $SB\chi^2$ statistic, the $SB\chi^2/df$ ratio (an acceptable cutoff value is 3.00), the robust comparative fit index (CFI; Bentler, 1990), the standardized root mean square residual (SRMR; Bentler, 1995), and the adjusted root mean squared error of approximation (RMSEA) for nonnormal conditions (see Nevitt & Hancock, 2000; Steiger & Lind, 1980). As recommended by Widaman and Thompson (2003), the CFI (i.e., an incremental fit index) was calculated with statistics from an "acceptable" independence null model, which is nested in the M_{scalar} analyses.[4] Values equal or superior to .90 for the CFI are considered to indicate a good fit, superior to .95 remaining ideal (Hu & Bentler, 1999).[5] Similarly, SRMR values equal to or smaller than .08 are considered acceptable (Hu & Bentler, 1999). Finally, values for RMSEA less than .08 are used as thresholds for not rejecting a model (Browne & Cudeck, 1992; see also Hu & Bentler, 1999).

In addition to these overall fit indexes, two types of CFIs were also used to statistically evaluate the difference between two nested models (see Bollen, 1989). First, the study used the scaled difference chi-square test ($\Delta SB\chi^2$; Satorra, 2000; Satorra & Bentler, 2001). Although it is not usually reported, the $\Delta SB\chi^2/\Delta df$ ratio to compare model differences for comparative purposes was also used. Second, as recommended by Cheung and Rensvold (2002), the study also examined the changes in the CFIs (i.e., the difference in the CFIs between two nested models or ΔCFIs) when invariance constraints were added. A value of ΔCFI smaller than or equal to $-.010$ indicates that the invariance hypothesis should not be rejected. One of the advantages of the ΔCFI over the $\Delta SB\chi^2$ is that it is not as strongly affected by sample size and number of constraints. Moreover, although much is known about the efficacy of $SB\chi^2$ and $\Delta SB\chi^2$, no research (to the best of our knowledge) has evaluated such statistics for mean structures. Therefore, although the study examined scalar invariance tests, the preferred focus was on overall indexes as well as ΔCFI.

---

[3]The Satorra-Bentler rescaled estimation has been chosen because of the somewhat nonnormarlity of the motivational variables, especially the amotivation subscale's items. The Mardia's coefficients (for multivariate kurtosis) were 569.23 and 536.28 for girl and boy samples, respectively.

[4]More specifically, besides estimated variances and means for manifest variables (corresponding to the traditional independence null model), in an acceptable independence null model means are constrained to be equal across groups and waves. Greater detail can be found in Widaman and Thompson's (2003) *Psychological Methods* article.

[5]Due to the complexity of the measurement model, it is not reasonable to expect a CFI to attain the .95 cutoff traditionally sought (see Marsh, Hau, & Wen, 2004).

RESULTS

In the following sections, the results for tests of MI are presented first, followed by the results of invariance tests on the latent constructs (i.e., latent standard deviations, correlations, and latent means). In each section, the general test for invariance across both waves and gender (longitudinal cross-gender invariance) is examined first and then gender and waves as sources of invariance are distinguished. Second, "simple invariance tests," which are cross-gender invariance tests for each of the three waves (i.e., 8th, 9th, and 10th grades) and longitudinal invariance tests across gender (i.e., for boys and girls) are reported. Finally, results from "interactional" invariance tests are examined.

MI

Overall fit indexes indicate good configural invariance of the AMS across waves and gender (Model 1A). Although $SB\chi^2$ was statistically significant, the other indexes are satisfactory ($SB\chi^2/df = 1.33$; CFI = .94; SRMR = .056; RMSEA = .023). Table 1 presents overall and comparative fit indexes for measurement models depicted in Figure 1 (a and b).

A general 2 × 3 MI.   Overall fit indexes ($SB\chi^2/df = 1.34$; CFI = .93; SRMR = .059; RMSEA = .023), as well as CFIs ($\Delta SB\chi^2/\Delta df = 1.78$; /$\Delta CFI = -.004$), provided good support for metric invariance across both waves and gender (Model 2B). Concerning scalar invariance (Model 3C), overall fit indexes were similar to those for the baseline and the metric invariance model ($SB\chi^2/df = 1.44$; CFI = .93; SRMR = .060; RMSEA = .026) and ΔCFI was lower than −.01. However, $\Delta SB\chi^2$ values were particularly "large." This is likely due to the asymptomatic nature that makes $\Delta SB\chi^2$ statistics oversensitive (see Discussion section). The $\Delta\chi^2$ statistics were less sensitive. On the other hand, a $\Delta CFI = -.007$ provides support for a satisfactory scalar invariance.

Looking at gender and longitudinal invariance separately, results indicate that metric invariance was stronger across gender (Model 1B) than waves (Model 2A). Indeed, there was no significant difference in $SB\chi^2$ and CFIs with the addition of invariance constraints to loadings across gender ($\Delta SB\chi^2 = 61.59$; $p = .051$; $\Delta SB\chi^2/\Delta df = 1.37$; $\Delta CFI = -.001$). Concerning longitudinal invariance, although $\Delta SB\chi^2$ was significant, other fit indexes remained good (e.g., $\Delta SB\chi^2/\Delta df = 1.93$; $\Delta CFI = -.004$). A similar pattern was observed at the scalar level, such that comparative statistics were better for gender (Model 1C; $\Delta\chi^2/\Delta df = 3.46$; $\Delta SB\chi^2/\Delta df = 11.78$; $\Delta CFI = -.005$) than longitudinal (Model 3A; $\Delta\chi^2/\Delta df = 5.18$; $\Delta SB\chi^2/\Delta df = 111.09$; $\Delta CFI = -.004$) invariance tests (see Table 1).

TABLE 1
Overall and Comparative Fit Indexes for the Nested Models for the Testing of Invariance Across Gender and Time

	Invariance Level		Overall Fit Indexes					Model	Comparative Fit Indexes[b]				
Models[a]	Longitudinal	Gender	$SB\chi^2/df$	CFI	SRMR	RMSEA	[90% CI]	Comparison	$\Delta SB\chi^2$	Δdf	p	$\Delta SB\chi^2/\Delta df$	ΔCFI
Longitudinal cross-gender invariance													
2B	Metric	Metric	1.34	.93	.059	.023	[.021, .025]	1A vs. 2B	133.48	75	.000	1.78	−.004
3C	Scalar	Scalar	1.44	.93	.060	.026	[.024, .028]	2B vs. 3C	5766.91	75	.000	76.89	−.007
									(365.11)	(75)	(.000)	(4.87)	
4D	SD	SD	1.39	.92	.093	.025	[.023, .026]	2B vs. 4D	198.08	25	.000	7.92	−.009
5E	Correlation	Correlation	1.42	.92	.092	.026	[.024, .027]	4D vs. 5E	156.39	50	.000	3.13	−.008
5E′	Correlation	Correlation	1.37	.93	.072	.024	[.022, .026]	3C vs. 5E′	150.43	50	.000	3.01	−.007
6F	Means	Means	1.60	.90	.086	.031	[.029, .032]	3C vs. 6F	−424.64	25	—	—	−.028
									(562.36)	(25)	(.000)	(22.49)	
Gender invariance													
1B	Configural	Metric	1.33	.94	.057	.023	[.021, .024]	1A vs. 1B	61.59	45	.051	1.37	−.001
1C	Configural	Scalar	1.37	.93	.057	.024	[.022, .026]	1B vs. 1C	529.88	45	.000	11.78	−.005
									(155.49)	(45)	(.000)	(3.46)	
1D	Configural	SD	1.35	.93	.082	.023	[.022, .025]	1B vs. 1D	91.34	15	.000	6.09	−.004
1E	Configural	Correlation	1.36	.93	.078	.024	[.022, .025]	1D vs. 1E	65.60	30	.000	2.19	−.003
1E′	Configural	Correlation	1.34	.93	.064	.023	[.021, .025]	1C vs. 1E′	58.82	30	.001	1.96	−.002
1F	Configural	Means	1.39	.93	.059	.025	[.023, .026]	1C vs. 1F	1156.97	15	.000	77.13	−.005
									(97.28)	(15)	(.000)	(6.49)	

Gender invariance in 8th grade

1B–8th	Configural	Metric in 8th	1.33	.94	.056	.023	[.021, .024]	1A vs. 1B	12.12	15	.670	.81	.000
1C–8th	Configural	Scalar in 8th	1.33	.94	.056	.023	[.021, .024]	1B vs. 1C	36.72	15	.001	2.24	.000
									(25.45)	(15)	(.044)	(1.70)	
1D–8th	Configural	SD in 8th	1.34	.94	.062	.023	[.021, .025]	1B vs. 1D	38.96	5	.000	7.79	−.002
1E–8th	Configural	Corr. in 8th	1.34	.93	.062	.023	[.021, .025]	1D vs. 1E	32.90	10	.000	3.29	−.002
1E′–8th	Configural	Corr. in 8th	1.33	.94	.060	.023	[.021, .025]	1C vs. 1E′	26.33	10	.003	2.63	−.001
1F–8th	Configural	Means in 8th	1.35	.94	.057	.023	[.021, .025]	1C vs. 1F	−87.11	5	—	—	−.003
									(58.53)	(5)	(.000)	(11.71)	

Gender invariance in 9th grade

1B–9th	Configural	Metric in 9th	1.33	.94	.056	.023	[.021, .024]	1A vs. 1B	14.76	15	.469	.98	.000
1C–9th	Configural	Scalar in 9th	1.34	.93	.056	.023	[.021, .025]	1B vs. 1C	2497.13	15	.000	166.48	−.003
									(64.92)	(15)	(.000)	(4.33)	
1D–9th	Configural	SD in 9th	1.33	.94	.061	.023	[.021, .025]	1B vs. 1D	13.94	5	.016	2.79	−.001
1E–9th	Configural	Corr. in 9th	1.33	.94	.061	.023	[.021, .025]	1D vs. 1E	19.46	10	.035	1.95	−.001
1E′–9th	Configural	Corr. in 9th	1.33	.94	.058	.023	[.021, .024]	1C vs. 1E′	19.70	10	.032	1.97	−.001
1F–9th	Configural	Means in 9th	1.35	.94	.056	.023	[.021, .025]	1C vs. 1F	44.19	5	.000	8.84	−.001
									(19.40)	(5)	(.002)	(3.88)	

Gender invariance in 10th grade

1B–10th	Configural	Metric in 10th	1.33	.94	.057	.023	[.021, .025]	1A vs. 1B	31.58	15	.007	2.11	−.001
1C–10th	Configural	Scalar in 10th	1.35	.93	.057	.023	[.022, .025]	1B vs. 1C	488.30	15	.000	32.55	−.002
									(65.58)	(15)	(.000)	(4.37)	
1D–10th	Configural	SD in 10th	1.35	.93	.073	.023	[.021, .025]	1B vs. 1D	89.58	5	.000	17.92	−.003
1E–10th	Configural	Corr. in 10th	1.35	.93	.070	.023	[.021, .025]	1D vs. 1E	24.53	10	.006	2.45	−.001
1E′–10th	Configural	Corr. in 10th	1.34	.94	.059	.023	[.021, .025]	1C vs. 1E′	22.71	10	.012	2.27	−.001
1F–10th	Configural	Means in 10th	1.37	.93	.058	.024	[.022, .026]	1C vs. 1F	−74.84	5	—	—	−.004
									(64.97)	(5)	(.001)	(12.99)	

TABLE 1 *(Continued)*

Models[a]	Invariance Level		Overall Fit Indexes					Model Comparison	Comparative Fit Indexes[b]				
	Longitudinal	Gender	SB_χ^2/df	CFI	SRMR	RMSEA	[90% CI]		ΔSB_χ^2	Δdf	p	$\Delta SB_\chi^2/\Delta df$	ΔCFI
Longitudinal invariance													
2A	Metric	Configural	1.34	.93	.059	.023	[.021, .025]	1A vs. 2A	115.96	60	.000	1.93	−.004
3A	Scalar	Configural	1.42	.93	.060	.026	[.024, .027]	2A vs. 3A	6665.38 (310.69)	60 (60)	.000 (.000)	111.09 (5.18)	−.006
4A	SD	Configural	1.37	.93	.075	.024	[.022, .026]	2A vs. 4A	134.91	20	.000	6.75	−.006
5A	Correlation	Configural	1.39	.92	.079	.025	[.023, .026]	4A vs. 5A	108.84	40	.000	2.72	−.005
5A′	Correlation	Configural	1.36	.93	.067	.024	[.022, .026]	3A vs. 5A′	113.98	40	.000	2.85	−.005
6A	Means	Configural	1.58	.90	.079	.030	[.028, .032]	3A vs. 6A	−305.01 (526.46)	20 (20)	— (.000)	— (26.32)	−.027
Longitudinal invariance for girls (G)													
2A-G	Metric for girls	Configural	1.34	.93	.058	.023	[.021, .025]	1A vs. 2A	68.82	30	.000	2.29	−.003
3A-G	Scalar for girls	Configural	1.39	.93	.059	.025	[.023, .026]	2A vs. 3A	24185.03 (167.61)	30 (30)	.000 (.000)	806.17 (5.59)	−.005
4A-G	SD for girls	Configural	1.36	.93	.067	.024	[.022, .025]	2A vs. 4A	62.71	10	.000	6.27	−.004
5A-G	Corr. for girls	Configural	1.37	.93	.067	.024	[.022, .026]	4A vs. 5A	47.93	20	.000	2.40	−.002
5A′-G	Corr. for girls	Configural	1.35	.93	.060	.023	[.022, .025]	3A vs. 5A′	52.38	20	.000	2.62	−.002
6A-G	Means for girls	Configural	1.47	.92	.067	.027	[.025, .029]	3A vs. 6A	−150.33 (270.80)	10 (10)	— (.000)	— (27.08)	– 0.14

Longitudinal invariance for boys (B)

2A-B	Metric for boys	Configural	1.33	.94	.057	.023	[.021, .024]	1A vs. 2A	41.28	30	.082	1.38	−.001
3A-B	Scalar for boys	Configural	1.37	.93	.057	.024	[.022, .026]	2A vs. 3A	1624.22 (143.06)	30 (30)	.000 (.000)	54.14 (4.77)	−.003
4A-B	SD for boys	Configural	1.34	.93	.065	.023	[.028, .032]	2A vs. 4A	76.07	10	.000	7.61	−.003
5A-B	Corr. for boys	Configural	1.36	.93	.070	.024	[.022, .028]	4A vs. 5A	61.23	20	.000	3.06	−.003
5A′-B	Corr. for boys	Configural	1.34	.93	.063	.023	[.021, .025]	3A vs. 5A′	62.35	20	.000	3.12	−.003
6A-B	Means for boys	Configural	1.44	.92	.070	.026	[.025, .028]	3A vs. 6A	−152.52 (257.30)	10 (10)	— (.001)	— (25.73)	−.014

Note. Model 1A, $\chi^2 = 4843.49$, $SB\chi^2 = 4106.73$, $df = 3090$, CFI = .94. Δ = a difference between the comparison model and the tested model; $SB\chi^2$ = the Satorra-Bentler maximum likelihood chi-square statistic; df = degrees of freedom; p = the probability level; CFI = robust confirmatory fit index; SRMR = standardized root mean square residual; RMSEA = root mean squared error of approximation; SD = Standard Deviation.
[a] Nested models were described in Figure 1.
[b] Comparison were made with baseline model and then with precedent (nested) models. For the comparisons based on the means structure the $\Delta\chi^2$ values are in parenthesis.

Cross-gender MI across waves.   Although previous results clearly indicated a good level of metric and scalar invariance across gender, it could be interesting to examine whether such invariance was constant across grade levels. First, concerning metric invariance, the difference in $SB\chi^2$ was smaller (indicating a higher degree of invariance) in 8th grade ($\Delta SB\chi^2 = 12.12$, $p = .670$), than in 9th grade ($\Delta SB\chi^2 = 14.76$, $p = .467$), and than in 10th grade ($\Delta SB\chi^2 = 31.58$, $p = .007$). As students aged, more subtle "variances" appeared between boys and girls. Second, concerning scalar invariance, the examination of chi-square differences (rather than $SB\chi^2$ differences) indicated a somewhat similar pattern in which more subtle variance appeared in 9th grade ($\Delta\chi^2 = 64.92$, $p < .001$; $\Delta SB\chi^2 = 2497.13$, $p < .001$) and 10th grade ($\Delta\chi^2 = 65.58$, $p < .001$; $\Delta SB\chi^2 = 488.30$, $p < .001$) than in 8th grade ($\Delta\chi^2 = 25.45$, $p < .044$; $\Delta SB\chi^2 = 36.72$, $p < .001$). These variations across waves can also be observed in ΔCFIs.[6]

Longitudinal MI across gender.   Because longitudinal invariance was less strongly reached, it is particularly interesting to examine whether gender could moderate this result. Results in Table 1 clearly indicate that metric invariance hypothesis was more supported in the boy sample ($\Delta SB\chi^2 = 41.28$; $p = .08$; ΔCFI = $-.001$) than in the girl sample ($\Delta SB\chi^2 = 68.82$, $p < .001$; ΔCFI = $-.003$). This difference between boys and girls was also obtained in scalar invariance tests ($\Delta\chi^2 = 167.60$ vs. 143.06, $ps < .001$; $\Delta SB\chi^2 = 1624.22$ vs. 24185.03, $ps < .001$; ΔCFI = $-.003$ vs. $-.005$).

Interactional MI.   We examined the impact of invariance constraints across gender on longitudinal invariance tests, and vice versa. For instance, the probability of rejecting cross-gender metric invariance was lower when constraints for longitudinal metric invariance were previously added (Model 2A vs. Model 2B; $\Delta SB\chi^2 = 17.75$; $p = .276$; $\Delta SB\chi^2/\Delta df = 1.18$) than when they were not (Model 1A vs. Model 1B; $\Delta SB\chi^2 = 61.59$; $p = .051$; $\Delta SB\chi^2/\Delta df = 1.78$). Similarly, when examining cross-gender scalar invariance, ΔCFI was lower when constraints related to longitudinal scalar invariance were previously added (Model 3B vs. Model 3C; ΔCFI = $.000$) than when they were not (Model 1B vs. Model 1C; ΔCFI = $-.005$). The inverse pattern seemed to appear for longitudinal invariance when cross-gender invariance constraints were added. For instance, when examining metric invariance, $\Delta SB\chi^2/\Delta df$ ratio was higher when constraints related to cross-gender metric invariance were previously added (Model 1B vs. Model 2B; $\Delta SB\chi^2/\Delta df = 2.53$) than when they were not (Model 1A vs. Model 2A; $\Delta SB\chi^2/\Delta df = 1.93$).

---

[6]It should be noted that these comparisons of the fit differences (as well as the next ones) are not statistical and, thus, somewhat arbitrated (see Discussion section).

This last result can be explained by the stronger metric invariance across gender than waves. Indeed, because the tests for cross-gender and longitudinal invariance shared common constraints (30 exactly; see Method section), CFIs can be different depending on which model they are based, than on models that do not account for these confounding constraints or on models in which these common constraints were previously added. Therefore, researchers need to pay attention to which nested models they used to test FI. The presence of shared common parameters could influence the meaning of the FI tests.

Latent Standard Deviation Invariance (LSDI)

A general 2 × 3 LSDI.   Overall and comparative fit indexes indicate the existence of a few degrees of variance in latent motivational factors' standard deviations (Model 4D). Indeed, although RMSEAs were below cutoff criteria, SRMR was higher than .09 and ΔCFI was very close to –.01 cutoff criteria. When distinguishing longitudinal and cross-gender invariance, fit indexes were more satisfactory but still revealed some variance. For instance, ΔSBχ^2/Δdf ratios were higher than 6.00 (Models 1D and 4A).

Cross-gender LSDI across waves.   As observed with scalar invariance, it seemed that cross-gender invariance was stronger in 9th grade than in 8th and 10th grade, ΔSBχ^2/Δdf ratio in 10th grade being particularly high (i.e., 17.92). The examination of standard deviation parameters for each motivational construct (Table 2) revealed that the lower degree of invariance in 8th grade was due to a lower standard deviation for external regulation for boys than girls ($\Delta\sigma = -.27$), and a higher amotivation's standard deviation for boys than girls ($\Delta\sigma = +.35$). In 10th grade, differences were mainly found for the amotivation's standard deviation, which was significantly lower for girls (i.e., $\sigma = .76$) than boys (i.e., $\sigma = 1.24$), $\Delta\chi^2 = 20.04$, $p < .001$. Although this stronger homogeneity in girls' reported amotivation than boys' was also observed in 8th and 9th grade, differences between standard deviations were less important (in 8th grade $\Delta\sigma = 1.35–1.00 = .35$; in 9th grade $\Delta\sigma = 1.49–1.20 = .29$; in 10th grade $\Delta\sigma = 1.24–0.76 = .48$).

Longitudinal LSDI across gender.   Longitudinal invariance of standard deviation was somewhat stronger for girls than boys (e.g., ΔSBχ^2/Δdf = 6.27 vs. 7.61), but this difference did not appear in ΔCFI. Examining relative standard deviations (Table 2) revealed that the longitudinal variations were present mainly in identified regulation and amotivation. For instance, among girls, standard deviations for identified regulation and amotivation seemed to increase from 8th to 9th grade and then decrease from 9th to 10th grade. A similar pattern can be observed for boys' amotivation, but boys remained heterogeneous in their identified regulation.

TABLE 2
Latent Standard Deviation for Each Motivational Construct

	Latent Standard Deviation		
	8th Grade	9th Grade	10th Grade
Intrinsic motivation			
Girls	1.00[a]	1.09	1.05
Boys	1.00	1.07	1.03
Identified regulation			
Girls	1.00[a]	1.30	1.14
Boys	0.98	1.44	1.38
Introjected regulation			
Girls	1.00[a]	1.15	1.08
Boys	1.15	1.15	1.16
External regulation			
Girls	1.00[a]	0.91	0.97
Boys	0.73	1.01	0.94
Amotivation			
Girls	1.00[a]	1.20	0.76
Boys	1.35	1.49	1.24

Note. Loadings are constrained to be equal across gender and waves.
[a] Fixed parameters.

Latent Correlation Invariance (LCI)

Because of the existence of a slightly variation at the standard deviation level, it could be preferable to examine LCI when latent standard deviation parameters were not constrained across gender and time (Model 5E′). Therefore, the fit indexes for models with and without standard deviation constraints (i.e., Model 5E and Model 5E′, respectively) are reported in Table 1.

A general 2 × 3 LCI. Overall fit indexes were satisfactory (Model 5E′: CFI = .93; RMSEA = .024; SRMR = .072). On the other hand, CFIs indicate the existence of a few degrees of variance in intercorrelations. Although ΔCFI was –.008 (i.e., below the –.01 cutoff but higher than the other ΔCFIs), ΔSBχ^2/Δdf ratio was higher than 3.00. However, when distinguishing longitudinal and cross-gender invariance, fit indexes were more satisfactory, ΔSBχ^2/Δdf ratios being lower than 3.00.

Cross-gender LCI across waves. The examination of cross-gender invariance in correlations revealed gender differences in correlations among motivational constructs mainly in 8th grade (ΔSBχ^2/Δdf = 2.63). For instance, correlations between amotivation and the other constructs were higher for girls (from –.06

to –.54) than for boys (from .00 to –.24) (see Table 3). When correlation invariance was tested in 8th grade including constraints on correlations other than those involving amotivation, $\Delta SB\chi^2$ was not statistically significant, $\Delta SB\chi^2(\Delta df = 6) =$ 16.37, $p = .012$. These differences were also present in 9th and 10th grade but were not as large.

Longitudinal LCI across gender.   As observed for standard deviations, girls' correlations were more longitudinally invariant than boys' ($\Delta SB\chi^2/\Delta df =$ 2.62 vs. 3.12). Examining correlation patterns (Table 3), it was observed that the larger differences in the boy sample were due to correlations between amotivation and the other constructs, especially between 9th and 10th grade. When "9th-to-10th-grade" constraints on correlations between amotivation and the other constructs were relaxed, $\Delta SB\chi^2$ was statistically nonsignificant, $\Delta SB\chi^2(\Delta df = 16)$ = 38.20, $p = .001$.

Latent Mean Invariance

In this section, the results from latent mean invariance tests (Table 1) are described first. Due to noninterpretable $\Delta SB\chi^2$ statistics (see Discussion section), overall fit indexes as well as $\Delta CFIs$ and $\Delta\chi^2$ statistics are focused on. Although it was not the purpose of this article, interpreted observed differences are then interpreted. Table 4 reports latent correlations when loadings and intercepts were constrained to be equal across gender and waves.

Invariance tests.   Results from longitudinal cross-gender invariance tests indicated differences in latent means ($\Delta CFI = -.028$; $\Delta SB\chi^2(\Delta df = 25) = 562.36$, $p <$ .001). Distinguishing longitudinal and gender effects, it was observed that latent mean differences were particularly significant across waves ($\Delta CFI = -.027$; $\Delta SB\chi^2(\Delta df = 20) = 526.46$, $p < .001$) and less significant across gender ($\Delta CFI =$ –.005; $\Delta SB\chi^2(\Delta df = 15) = 97.28$, $p < .001$). More specific invariance tests revealed that latent mean differences between girls and boys were higher in 8th and 10th grades ($\Delta CFIs = -.003$ and $-.004$; $\Delta SB\chi^2(\Delta df = 5) = 58.53$ and 64.97, $ps < .001$) than in 9th grade ($\Delta CFI = -.001$; $\Delta SB\chi^2(\Delta df = 5) = 19.40$, $p = .002$). However, it should be noted that although intercept invariance was less evident in 9th grade than in others waves (see previous), intercepts were constrained to be equal across samples. Therefore, any gender differences in raw scores that could be observed in 9th grade should be therefore mainly interpreted in terms of an intercept difference rather than a construct difference. This was less true in 8th and 10th grade. Longitudinal invariance was equivalent in both girl and boy samples ($\Delta CFIs = -.014$).

Latent means interpretation.   Table 4 revealed the existence of various 2 × 3 latent mean patterns depending on the motivational constructs. For instance,

TABLE 3
Latent Correlations Among Motivational Constructs for Girls (Upper Diagonal) and for Boys (Lower Diagonal)

	8th Grade					9th Grade					10th Grade				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Intrinsic motivation (1)	—	.80	.63	.30	−.59	—	.72	.72	.46	−.47	—	.81	.65	.17	−.67
		(.78)	(.56)	(.20)	(−.54)		(.73)	(.72)	(.49)	(−.47)		(.78)	(.64)	(.16)	(.64)
Identified regulation (2)	.70	—	.68	.73	−.55	.71	—	.59	.90	−.45	.71	—	.65	.57	−.67
	(.70)		(.64)	(.72)	(−.49)	(.74)		(.60)	(.92)	(−.45)	(.73)		(.65)	(.60)	(−.61)
Introjected regulation (3)	.78	.59	—	.57	−.38	.78	.57	—	.54	−.35	.62	.51	—	.49	−.35
	(.77)	(.58)		(.54)	(−.27)	(.78)	(.58)		(.60)	(−.34)	(.61)	(.50)		(.49)	(.33)
External regulation (4)	.39	.83	.42	—	−.15	.47	.89	.45	—	−.18	.23	.71	.29	—	−.03
	(.31)	(.79)	(.36)		(−.06)	(.53)	(.89)	(.49)		(−.21)	(.29)	(.73)	(.33)		(−.04)
Amotivation (5)	−.21	−.20	−.09	−.01	—	−.26	−.47	−.07	−.26	—	−.47	−.69	−.23	−.25	—
	(−.24)	(−.26)	(−.10)	(.00)		(−.32)	(−.53)	(−.10)	(−.31)		(−.51)	(−.72)	(−.22)	(−.30)	

Note. Loadings and latent standard deviations are constrained to be equal across gender and waves. In parenthesis only loadings are constrained to be equal across gender and waves.

TABLE 4
Latent Means For Each Motivational Construct

	Latent Means		
	8th Grade	*9th Grade*	*10th Grade*
Intrinsic motivation			
Girls	.00[a]	−.52	−.88
Boys	−.24	−.63	−1.10
Identified regulation			
Girls	.00[a]	−.39	−.66
Boys	.04	−.69	−1.09
Introjected regulation			
Girls	.00[a]	−.33	−.78
Boys	−.28	−.53	−1.17
External regulation			
Girls	.00[a]	.17	−.23
Boys	.44	.15	−.01
Amotivation			
Girls	.00[a]	.45	.02
Boys	.29	.81	.49

Note.   Loadings and intercepts are constrained to be equal across gender and waves.
[a]Fixed parameters.

latent means for intrinsic motivation, identified regulation, and introjected regulation decreased from 8th to 10th grade and were higher for girls than boys. However, interactions can be observed only in ID latent means where gender differences were absent in 8th grade but significant in 9th grade and largest in 10th grade. Concerning external regulation, boys' latent means consistently decreased from 8th to 10th grade, whereas girls' external regulation increased from 8th to 9th grade and then decreased from 9th to 10th grade. The combination of these two patterns resulted in gender differences (boys' being higher than girls') only in 8th and 10th grade. Finally, both girls' and boys' amotivation latent means increased during the first period (8th to 9th grade) and then decreased in the second period (9th to 10th grade).

DISCUSSION AND CONCLUSION

The purpose of this study was to evaluate the longitudinal cross-gender FI of the AMS (Vallerand et al., 1989, 1992, 1993). A second related purpose was to present an integrated analytical strategy to test both longitudinal (across 3 years) and cross-gender FI of a multidimensional scale in one set of nested models. This 2 × 3 design allowed us to test specific hypotheses concerning cross-sample FI for each

wave, as well as the comparison of longitudinal FI in the boy and girl samples. These results have also implications for FI research.

Implications for AMS Validation

These results provide support for longitudinal cross-gender metric invariance, as well as satisfactory support for scalar invariance of the AMS. Therefore, we recommend using this scale to test specific hypotheses about gender differences as well as motivational development. More specific results revealed that cross-gender invariance was slightly stronger than longitudinal invariance. Furthermore, subtle cross-gender variance appeared in the measure as students became older. Because longitudinal invariance seemed to be stronger for boys than for girls, one can conclude that girls subjectively redefined the motivational construct, whereas boys seemed to be constant. We propose two types of explanations for this finding.

First, the difference found in the redefinition of motivational constructs may be due to developmental differences between girls and boys. Indeed, the period from 8th to 10th grades coincides with early adolescence and girls usually reach pubertal maturity earlier than boys. Because of the complexity of the constructs, measurement characteristics may change during childhood, perhaps differently for girls and boys. An alternative explanation concerns familiarity with the motivational questionnaire. Girls might readjust their responses, whereas boys tended to remain constant in their definition of the constructs. Fortunately, FI tests were still satisfactory and allow further tests at the latent construct level.

Other interesting results are related to the latent constructs. For instance, whereas girls were more heterogeneous in amotivation than boys, the reverse was observed for external regulation. Furthermore, the 9th grade appears to be developmentally significant for girls because they were more heterogeneous in identified regulation and amotivation only during this wave compared to boys. This last finding can be explained by the previous results obtained with metric invariance. Although we obtained a satisfactory degree of metric invariance, the similarity in pattern that we observed in metric and standard deviations invariance tests could reveal the influence of the former on the later. However, once again an alternative explanation concerns maturity. Perhaps girls developed a relatively higher degree of maturity in 9th grade than boys. Further examination of data in 11th and 12th grade will be needed to test this hypothesis and to verify if a higher degree in boys' standard deviation will be observed later than girls.

Implications for Testing FI Within Structural Equation Modeling Framework

The second purpose of this study was to demonstrate how to test both longitudinal and cross-gender factorial invariance of a multidimensional scale in one set of

nested models. Notably, five points should be emphasized. First, this analytical strategy distinguished measurement and latent construct components of the factorial model (see e.g., T.D. Little, 1997).

Second, a specific approach to solve identification problems in covariance and mean structures was used. Specifically, this study proposed, first, to fix factor variance and means in one referent group, and then, to constrain the loading and intercept of one item to be equal across samples and waves. In addition to the fact that this approach could be equivalent to others in testing global (rather than partial) MI, the proposed approach is particularly pertinent for testing latent construct invariance in a 2 × 3 design.

Third, this study also represented an opportunity to examine the efficacy of various comparative indexes for comparing two nested models. Concerning the difference between two $SB\chi^2$ statistics, although the test seemed appropriate for testing restrictions in covariance structure, $\Delta SB\chi^2$ values were noninterpretable when restrictions were on mean structure (i.e., intercepts and latent means). Satorra and Bentler (2001) suggested that $\Delta SB\chi^2$ values (especially difference in scaling corrections) may turn out to be large or negative in particularly extreme cases, such as when sample size is small or when a less restricted model is "too deviant from the true model" (p. 511). Although these two extreme cases could pertain to this study, results clearly indicated that the analysis of a mean structure (in addition to a covariance structure) could also provide a source of explanation. Further research is needed to examine the behavior of $\Delta SB\chi^2$ and the conditions when it can be applied to test invariance in a mean structure.

Fourth, this study also used the $\Delta SB\chi^2/\Delta df$ ratio as a CFI. Like the well-known χ^2/df ratio, the ratio between the chi-square difference and the difference of degrees of freedom is particularly useful to examine how two nested models differ. Moreover, the ratio allowed for a comparison among invariance tests, especially when Δdfs varied. Because the $\Delta SB\chi^2/\Delta df$ ratio is relatively new in invariance testing, further examination in simulation studies is needed. Concerning the ΔCFIs, although only Cheung and Rensvold (1999) have appropriately studied the change in CFI as an indicator of (in)variance and further studies are needed, this test discriminated well among the various degrees of variance that were found in the study. In particular, the ΔCFI was more appropriate for invariance testing in mean structure than the $SB\chi^2$ statistics. Finally, it is important to stress the importance of the use of various overall and comparative fit indexes to deduce the invariance of any parameters across samples, waves, or both.

Finally, in examining cross-gender invariance across waves and longitudinal invariance across gender, the relative fit loss (i.e., $\Delta SB\chi^2$, $\Delta SB\chi^2/\Delta df$ ratio, and ΔCFI) was compared. For example, the study showed that the gender difference in $SB\chi^2$ (for cross-gender metric invariance) was smaller in 8th grade than in 9th grade and than in 10th grade (see section Cross-Gender MI across waves). However, it is important to note that these comparisons are not statistical comparisons

and, thus, are somewhat arbitrated. Further research is needed to find a statistical method for comparing two invariance tests (e.g., two $\Delta\chi^2$) and to determine what constitutes a significant difference.

To conclude, we believe that this article and its associated results provide strong support for the use of the AMS to compare the motivation of girls and boys, as well as to compare across grade levels. Finally, we hope that the reader will be convinced of the usefulness of the integrated analytical strategy that we propose.

ACKNOWLEDGMENTS

REFERENCES

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246.

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research, 21,* 230–258.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25,* 1–27.

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

Deci, E. L. (1975). *Intrinsic motivation.* New York: Plenum.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York: Plenum.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11,* 227–268.

Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist, 26,* 325–346.

Guay, F., & Vallerand, R. J. (1997). Social context, student's motivation, and academic achievement: Toward a process model. *Social Psychology of Education, 1,* 211–233.

Harter, S. (1981). A new self-orientation scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developmental Psychology, 17,* 300–312.

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

Jamshidian, M., & Bentler, P. M. (1999). ML estimation of mean and covariance structures with missing data using complete data routines. *Journal of Educational and Behavioral Statistics, 24,* 21–41.

Kim, K. H., & Bentler, P. M. (2002). Tests of homogeneity of means and covariance matrices for multivariate incomplete data. *Psychometrika, 67,* 609–624.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

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

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11,* 320–341.

McArdle, J. J., & Cattell, R. B. (1994). Structural equation models of factorial invariance in parallel proportional profiles and oblique confactor problems. *Multivariate Behavioral Research, 29,* 63–113.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525–543.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 52,* 431–462.

Nevitt, J., & Hancock, G. R. (2000). Improving the root mean square error of approximation for nonnormal conditions in structural equation modeling. *The Journal of Experimental Education, 68,* 251–268.

Pelletier, L. G. (2004). *Academic motivation and dropout in Ottawa elementary and high school (2000–2004).* Report submitted to Le Conseil des Écoles Catholiques de Langue Française de l'Est de l'Ontario, University of Ottawa, Ontario, Canada.

Pelletier, L. G., Séguin-Lévesque, C., & Legault, L. (2002). Pressure from above and pressure from below as determinants of teachers' motivation and teaching behaviors. *Journal of Educational Psychology, 94,* 186–196.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114,* 552–566.

Satorra, A. (2000). Scaled and adjusted restricted tests in multisample analysis of moment structures. In D. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis: A Festschrift for Heinz Neudecker* (pp. 233–247). Dordrecht, The Netherlands: Kluwer.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *1988 Proceedings of the American Statistical Association,* 308–313.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

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

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

Steiger, J. H., & Lind, J. C. (1980, June). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Thompson, B. (2003). *Score reliability: Contemporary thinking on reliability issues.* Thousand Oaks, CA: Sage.

Thurstone, L. L. (1947). *Multiple-factor analysis: A development and expansion of the vectors of mind.* Chicago: University of Chicago Press.

Vallerand, R. J., Blais, M. R., Brière, N. M., & Pelletier, L. G. (1989). Construction et validation de l'échelle de motivation en éducation (ÉMÉ) [Construction and validation of the Échelle de Motivation en Éducation—ÉMÉ]. *Canadian Journal of Behavioural Science, 21,* 323–349.

Vallerand, R. J., Fortier, M. S., & Guay, F. (1997). Self-determination and persistence in a real-life setting: Toward a motivational model of high school dropout. *Journal of Personality and Social Psychology, 72,* 1161–1176.

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Brière, N. M., Senécal, C. B., & Vallières, É. F. (1992). The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement, 52,* 1003–1017.

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Brière, N. M., Senécal, C. B., & Vallières, É. F. (1993). On the assessment of intrinsic, extrinsic, and amotivation in education: Evidence on the concurrent and construct validity of the academic motivational scale. *Educational and Psychological Measurement, 53,* 159–172.

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.

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods, 8,* 16–37.