Factor Mixture Models: Mixture Modeling as a Tool for Studying Measurement Invariance

Abigail R. Lau
Dena A. Pastor

James Madison University
Center for Assessment and Research Studies

Paper presented at the annual conference of the Northeastern Educational Research Association, Rocky Hill, CT, October 2007.

Abstract

Factor mixture modeling (FMM) is latent variable technique that can identify unobserved subgroups within a sample across which factor model parameters vary. If no such unobserved groups are identified, FMM results provide strong evidence that a single structure applies. In the current study FMM is applied to the Motive to Avoid Failure Scale (MAF; Hagtvet & Benson, 1997), a measure of a maladaptive achievement motivation.

Factor Mixture Models: Mixture Modeling as a Tool for Studying Measurement Invariance

Imagine a student who feels dread, or even fear, when faced with a challenging academic task. Motive to Avoid Failure (MF), describes this tendency of an individual to anticipate shame, humiliation, or other negative affect when the outcome of a task is uncertain (Atkinson & Litwin, 1960; Atkinson, 1957). Students with high MF are expected to exhibit two key characteristics. First, when asked about their expectations for their future performance, they are likely to predict that they will do much better or much worse than is plausible given their actual ability level (referred to as defensiveness). Second, when given the choice, they are likely to select tasks that are either far below or far above their ability level, rather than tasks at their appropriate level (Ceranski, Teevan, & Kalle, 1979). These two characteristics have been shown to cause students to adopt avoidance goal orientations that are associated with lower achievement outcomes (Conroy & Elliot, 2004; Diseth & Martinsen, 2003; Fairchild, Horst, Finney, & Barron, 2005). Further, students high in MF typically experience higher rates of failures and lower academic self-efficacy (Covington, 1984). Given these important implications for students, researchers seek to understand the correlates and antecedents of MF (Conroy & Elliot, 2004; Minnaert, 1999; Minnaert, 2003).

Measurement of MF

Conceptually, MF is analogous to test anxiety, and differs from it only in the terms of the contexts that are applicable (Elliot & Church, 1997; Hagtvet & Benson, 1997; Minnaert, 1999; Minnaert, 2003). Test anxiety is a specific type of MF that occurs in testing situations, whereas more general MF also applies to other achievement situations, such as sports and learning new skills. Because of this similarity, MF was originally measured using the same instruments as test anxiety (Atkinson, 1957; Atkinson & Litwin, 1960). However, scales have been developed to specifically measure MF. For example, the Achievement Motives Scale (AMS; Nygárd & Gjesme, 1973) includes a 12-item subscale designed to assess general MF. The Motive to Avoid

Table 1

Motive to Avoid Failure Scale Items

Order	Item
1	I dislike work that I’m not sure I can manage.
2	I’m afraid of failing in situations where the outcome is uncertain.
3	I dislike doing things that seem somewhat difficult.
4	Just thinking about working on new, somewhat difficult tasks makes me feel uneasy.
5	I dislike working in situations if I am uncertain how well I will do.
6	I am afraid of failing when I am given a task that I am uncertain that I can solve.

Failure Scale (MAF; Hagtvet & Benson, 1997) is a brief measure that targets only MF and can be used in a wide variety of performance settings. The MAF scale includes six items from the AMS subscale that closely align with the definition of MF. Specifically, items were selected for the MAF to reflect two characteristics of MF: a) anticipation of negative affects and b) uncertain outcomes.

The six MAF items are shown in Table 1. Students respond to each item using a four point scale where 1 is 'strongly disagree,' 2 is 'disagree', 3 is 'agree,' and 4 is 'strongly agree.' Previous studies have not reported internal-reliability coefficients for the scale's scores. *Internal Structure of the MAF*

Finding that items relate to each other in theoretically aligned ways is an important piece of validity evidence for scores. Some internal validity evidence for MAF scores exists in that the items have shown adequate fit to a one-factor confirmatory factor model (Hagtvet & Benson, 1997; Young, Finney, & Lau, 2005). However, this finding alone is insufficient internal validity evidence. When a factor model is fit to data, an assumption is made that the data were sampled from a single homogeneous population. In reality, populations can have multiple subpopulations, which are mixed together in the observed sample. Furthermore, factor model parameters can vary across these subpopulations, which impacts the comparability of scores across those groups. Thus, in order to be confident about the comparability of MAF scores, measurement invariance studies are needed.

Commonly, measurement invariance studies are conducted using multiple-groups structural equation models or item-response theory. A common approach with these methods is to estimate the measurement model parameters for each subpopulation separately and then compare the resulting estimates. One limitation of these techniques is that the subpopulations must be known a priori, so measurement invariance can only be investigated across observed groups. An alternative technique, known as factor mixture modeling (FMM), can be used to examine if unobserved subpopulations exist that differ in their measurement model parameters. Another difference between FMM and multiple-groups analyses is the order in which research questions are asked. When conducting a multiple groups analysis researchers first consider what known groups might differ in their measurement parameters for the scale. Then, researchers empirically investigate whether the measure is invariance across those known groups. When conducting an FMM researchers first investigate whether unobserved groups exist for which the measurement model parameters differ, and then attempt to identify the characteristics of the unobserved groups.

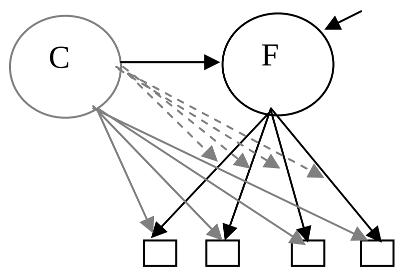
Approaching measurement invariance studies though FMM is less common and can be computationally complex. However, FMMs have two great advantages: a) they can detect subpopulations driven by unobservable variables or a combination of variables, and b) they can efficiently determine whether subpopulations exist, to obviate the need to test each type of subpopulation. Thus, FMM is an efficient approach to adding to the internal validity evidence for MAF scores.

Factor Mixture Models (FMMs)

A FMM is considered a hybrid latent model because it contains both a continuous latent variable found in a traditional factor model, and a categorical latent variable found in a traditional latent class or latent profile model. Specifically, in a one-factor FMM, the factor is the latent construct the instrument is intended to measure, and the levels of the categorical latent variable represent different latent subpopulations or classes. FMMs merge these two common types of latent models by relaxing the restrictions implied by either type. For example, in the latent class or latent profile model, relationships among observed variables are assumed to be completely explained by a student's latent class membership. In the factor model, relationships among the observed variables are assumed to be completely explained by the factor. However, in FMMs, variance in the item is assumed to be completely explained by the combination of the factor and the latent class. In other words, students within the same class are allowed to vary in their level of the factor, and students with the same factor score could be from different classes.

Figure 2 is a diagram of a FMM. The latent class variable, representing the latent subpopulations, is allowed to influence the measurement model parameters (intercepts and slopes) and the factor variances[1]. With this model, the factor model parameters are allowed to vary depending on the subpopulation. When measurement model parameters vary across the levels of the class variable, the measurement model is said to be non-invariant across those subpopulations.

Figure 2. Factor mixture model diagram.



Note. C is a categorical latent variable; F is the continuous latent variable (factor).

A FMM with a one level models a situation in which all cases are from the same population, a model identifical to the traditional factor model. A FMM with two classes represents a situation where there are two latent subpopulations, allowed to differ from one another in their factor model parameters. Measurement invariance for a scale is examined by posing multiple FMMs that vary in the number of levels specified for the latent categorical variable. If the model with only a single level of the class variable fits just as well as a model with multiple levels of the class variable, there is evidence that the measurement model is invariant.

Purpose of the Study

The purpose of the current study is to better understand the internal structure of the MAF scale using factor mixture modeling. Specifically, our research questions are: 1) Do latent subpopulations exist? 2) If latent subpopulations exist, which factor model parameters vary across the subpopulations? 2) If latent subpopulations exist, what are the characteristics of the subpopulations? The first research question will be examined by comparing the fit of various FMMs, each differing in the number of classes (1 to 5) specified. If an FMM with more than one class fits the data, the second research question will be examined by comparing the factor model

[1] Factor means are constrained to zero in each class for identification purposes.

parameters of the various classes to one another. If multiple classes exist, an attempt will be made to understand the source of the measurement invariance by examining the relationships of class membership with other variables.

Methods

Participants and Procedures

Students at a mid-sized southeastern university completed the MAF items as part of a battery of cognitive and developmental tests that were administered under standardized conditions for the purpose of university program assessment. The sample used in this study is a concatenation of data collected in the spring of 2002 (N = 2,261) and the spring of 2003 (N = 1,856). In each sample, students were approximately 64% female, 85% Caucasian, and in their sophomore (75%) or junior (25%) year of college. These demographic percentages roughly mirror those of the university population. Participation in the assessments was mandatory for all students who had earned between 40-75 credit hours. Students who met the criteria, and did not attend the assessment, had a hold placed on their student record until they were able to meet the requirement. Although participation is required, data collected are only used in aggregate, so individual scores are not reported and have no consequences for students. Thus, for some students, motivation to attend to the items is low. Only students with complete data on the MAF scale were included in the sample.

External Variables

An external variable is a variable that was not used to classify students into latent classes. If multiple latent classes were found, the external variables were compared across classes in an attempt to characterize the unobserved subpopulations.

Achievement Goal Orientation. The modified version (Finny, Pieper, & Barron, 2004) of Elliot and McGregor's (2004) Achievement Goal Orientation Questionnaire was used to measure students levels of four achievement goal orientations in their courses that semester. Students responded to 12 statements, with three items used to measure each of the four goal orientations, on a scale ranging from 1 (not at all true of me) to 7 (very true of me). Higher scores on each of the subscales indicate that the student had higher levels of that goal orientation in their courses that semester. The goal orientations were: mastery-approach, performance-approach, performance-avoidance, and mastery-avoidance. Mastery-approach goals are when students seek to master the material, whereas performance approach goals are when students seek to perform better than others. Mastery-avoidance goals are when students focus on avoiding misunderstanding the material, and performance-avoidance goals occur when student focus on avoiding performing poorly compared to others.

Additionally, four items were included that measure work avoidance (Harackiewicz et al., 1997). An example work avoidance item is "I want to do as little work as possible this semester."

Theories of Intelligence (TOI). The TOI (Dweck, Chiu, & Hong, 1995) measures students' implicit attitudes and beliefs about malleability of their intelligence. An example item is "You have a certain amount of intelligence, and you can't really do much to change it." The scale used here contains four items which students respond to using a scale of one to six, with one indicating 'strongly disagree' and six indicating 'strongly agree'. Higher scores indicate that a student believes their intelligence is malleable.

Examinee Motivation. Students self-reported the amount of effort they exerted on the instruments using the Student Opinion Survey (SOS; Sundre & Moore, 2002; Wolf & Smith, 1995). The SOS consists of five Likert items, on a 1 to 5 scale, from the 'strongly disagree' to 'strongly agree.' Higher scores on the scale indicate higher levels of effort put forth on the test.

Models

A one-factor model of MAF was specified with one to five classes. For each model the item intercepts, factor loadings, and residuals were allowed to vary across classes. The factor means for all classes were constrained to zero for identification purposes, whereas the factor variances were allowed to vary across classes. The slope of item 1 was constrained to one to set the scale of the factor.

Model Comparison

In order to determine whether a *k* or *k-1* class solution was more empirically supported, several indices were used to compare the models. The most basic index is the log-likelihood (LL) value that results from maximum likelihood estimation. The LL gives an indication how likely the sample data is given the parameter estimates. Higher LL values indicate better model fit. More complex models will always yield higher LLs than more parsimonious models.

Other fit indices based on the LL values can be used to compare models. The Bayesian information criteria (BIC; Schwartz, 1978) is one such index formed using the following equation: BIC = $-2\ \text{LL} + p \ln N$, where p is the number of parameters and N is the sample size. The sample-size adjusted BIC (SSA-BIC), in which the value for N is $((N+2)/24)$, was also used. Lower values of the BIC and SSA-BIC indicate better model fit.

Data analysis

Mplus version 4.1 software (Muthén & Muthén, 2005) was used to estimate the factor mixture models with maximum likelihood estimation using the EM algorithm. To prevent interpreting a solution based on local maxima, each model was estimated 100 times, each time start values for the parameters were randomly generated. The resulting ten solutions with the highest LL values went through final optimizations. If all ten solutions did not converge on the same LL value, the similarity if the parameter estimates obtained for each solution was examined. Model parameters were only interpreted if the ten final solutions had approximately equal parameter estimates.

For each interpreted model solution, Mplus provided the probability that each student belonged in each class. In addition, Mplus assigns each student to a class based on which class they are most likely to belong. After students have been assigned to a class, the characteristics of the student in each class were examined to identify defining characteristics of the classes. To conduct these analyses student demographic information, including GPA, gender, ethnicity, and year in school was obtained.

Results

Research Question 1: Do latent subpopulations exist?

Primarily, we were interested in whether or not different subpopulations of students exist for which the factor model parameters are different. The presence of subpopulations would be indicated by empirical support for a two or more class solution over a one-class one-factor model. Table 2 presents the information indices for the one-class and two-class solutions; a stable solution for the three-class model could not be identified. All indices indicate that the two-class solution fit the data substantially better than the one-class solution. Thus, the answer to our first research question was yes, latent subpopulations exist. Specifically, the subgroups appear to each account for approximately half the population, with an estimated 52% of the population belonging to Class 1. Furthermore, the classes appear to be fairly clearly defined, since the average latent class probability was .87 for both classes.

Table 2

Model Comparison Indices for Class Models

Model	parameters	LL	AIC	BIC	SBIC
1 class	18	-23536.4	47108.86	47222.09	47164.9
2 class	37	-22625	45323.96	45556.72	45439.15
3 class*	44				

*Indices not reported because model failed to converge to a stable solution.

Research Question 2: If latent subpopulations exist, on which factor model parameters do they vary?

After uncovering latent subgroups, we were interested in understanding how the factor model parameters varied across the classes. Factor means across the classes could not be compared because they were constrained to zero across in each class for identification purposes. However, factor variances were estimated freely for each class. In the two-class solution, Class 1 factor variance was estimated to be .06, whereas Class 2 factor variance was estimated as .09. Thus, one of the differences across classes is the subpopulation Class 2 is more varied in their levels of MAF.

Table 3 presents the item parameters for each class and the absolute difference between the parameters. Figure 3 is a plot of the item intercepts for both classes which shows that the intercepts were higher for Class 2 across all items. Recall, that the model was specified such that the factor means of all classes were constrained to zero. Thus, any differences in factor means would be apparent by consistent differences in item intercepts across items, similar to what is observed in Figure 3. Thus, one interpretation of these intercept differences is that the items are functioning the same across subpopulations, but the subpopulations differ in their levels of the factor. However, the intercept differences between classes for items 1, 2 and 6 is somewhat

Table 3

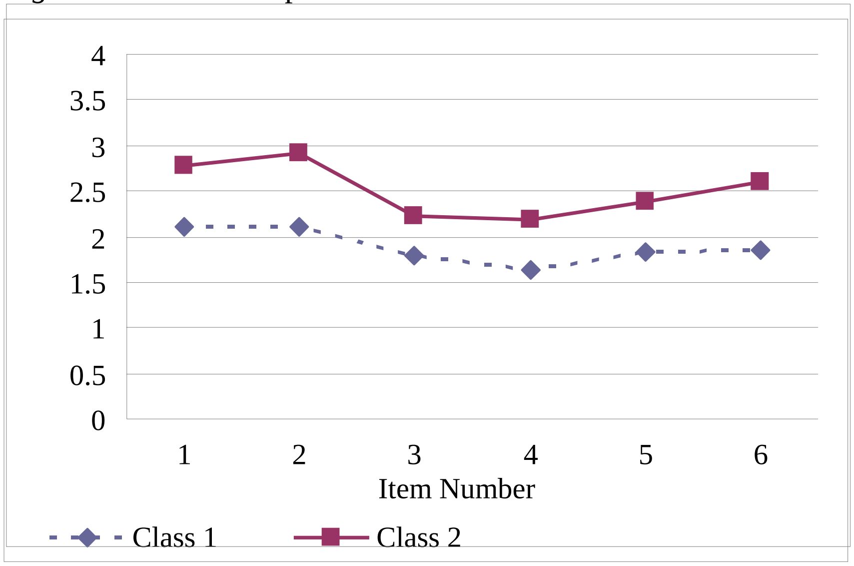
Factor Model Parameter Estimates for the Two-Class Solution

	Intercepts			Loadings*			Residuals*		
Item	Class 1	Class 2	Abs. Diff.	Class 1	Class 2	Abs. Diff.	Class 1	Class 2	Abs. Diff.
1	2.10	2.78	0.68	0.49	0.37	0.12	0.76	0.86	0.10
2	2.11	2.91	0.80	0.65	0.42	0.23	0.58	0.82	0.24
3	1.80	2.23	0.43	0.55	0.55	0.00	0.70	0.70	0.00
4	1.64	2.19	0.55	0.61	0.70	0.09	0.63	0.52	0.11
5	1.83	2.38	0.55	0.73	0.76	0.04	0.47	0.42	0.05
6	1.86	2.60	0.74	0.76	0.60	0.15	0.43	0.64	0.21

* Completely Standardized

Note. Abs. Diff. = Absolute difference between Class 1 and Class 2 estimates.

Figure 3. Item intercepts for the two latent classes.



larger than for items 3, 4 and 5. Thus, the pattern of intercepts for the two classes is roughly parallel, but not strictly parallel. Variances in the *pattern* of item intercepts does suggest that the measurement may not be invariant across classes.

The pattern of completely standardized factor loadings is somewhat differs across classes in a similar pattern to the item intercepts. Specifically, the factor loadings differ by a larger amount for items 1, 2 and 6 and by a smaller amount for items 3, 4, 5. It seems that items 1, 2 and 6 were less related to the factor being measured in Class 2 in comparison to Class 1.

Only a formal statistical test of nested models could determine whether these differences are large enough that measurement should be considered noninvariant across the classes.

However, just glancing at the differences helps shed light on what may be differentiating the classes. For example, there appear to be two sets of items (1, 2, 6 vs. 3, 4, 5) with the former set suggesting larger differences between the classes than the latter. Efforts to attain measurement invariance should first be focused on this set and particularly on items 2 and 6, which appear to be the most problematic.

Research Question 3: What are the characteristics of the subpopulations?

Each student is assigned to one or the other latent class based on which class they are most likely to belong. Thus, the characteristics of the subpopulations can be examined by considering the characteristics of the students assigned to each class. The proportion of students in each class for each gender, ethnicity, and year in college is reported in Table 4. The classes had approximately equal demographic profiles. We also examined the motivational profiles of the students assigned to each class. Specifically, Table 5 shows the mean differences across classes on achievement goals, examinee motivation, theories of intelligence, and work avoidance. The classes differed in all variables except for Performance Approach orientation,

Table 4

Demographic Composition of the Latent Classes

	Class 1		Class 2	
	N	%	*N*	%
Females	1,287	61.37	1,285	62.7
Males	810	38.63	705	37.3
White	1,805	86.08	1,593	84.29
Black	76	3.62	70	3.7
Hispanic	31	1.48	33	1.75
Asian	96	4.58	109	5.77
American Indian	5	0.24	2	0.11
Not Specified	84	4.01	83	4.39
Sophomores	1,576	75.15	1,413	74.76
Junior	521	24.85	476	25.19

which describes a student's interest in performing so that they look good relative to others. In general, students in Class 2 were higher in the less desirable motivation constructs and lower in the more adaptive orientations. This means that students in Class 2 were less likely to set goals to master the material, and more likely to set goals to avoid looking bad compared to others or to only due the minimum required. In addition, students in Class 2 had slightly lower GPAs, and believed their intelligence was fixed.

Class member comparisons added support to the idea that students in Class 2 are higher in overall MF level, since MF is expected to be associated with negative achievement outcomes. Thus, there is more evidence that the differences in item intercepts across classes were due to difference on the factor scores rather than differential item functioning. The observed and factor MAF scores also indicate that Class 2 is higher in MF on average. However, comparison of these scores is not necessarily trustworthy, because they may not mean the same thing across classes. In sum, class characteristic comparisons aided in interpretation of the item parameters, but did not reveal any distinctions between the classes that would help explain their theoretical meaning.

Table 5

Differences Between Classes in External Variables

	Class 1		Class 2		*d*
	M	*SD*	*M*	*SD*	
Mastery Avoidance	14.38	4.78	16.42	4.98	-0.42
Work Avoidance	12.75	3.47	13.99	3.95	-0.33
Performance Avoidance	12.15	4.06	13.39	4.02	-0.31
Theories of Intelligence	17.96	4.50	17.00	4.91	0.21
Mastery Approach	22.60	3.06	21.62	3.43	0.24
GPA	2.91	.620	2.84	0.68	0.12
Examinee Effort	17.79	3.90	17.27	4.33	0.13
Performance Approach	15.47	4.07	15.46	4.16	0.00

Discussion

The purpose of this study was to examine whether the assumption of population homogeneity holds for the MAF scale. We found evidence for two latent classes for which factor model parameters vary, indicating that the assumption of population homogeneity did not hold. We found evidence that the classes differ in their factor means and variances as well as item parameters for items 1, 2, and 6. We did not find evidence that the latent classes were related to observed population membership such as gender, ethnicity, or college-year. However, we did find that motivational profiles of students in Class 2 were consistent having a higher MF level.

Implications

Based on the current study alone, we are not able to draw conclusions about what is driving the latent classes. The latent classes could have emerged for a multitude of reasons, and may have direct or indirect meaning. A direct interpretation of the latent groups would be that the groups represent some theoretically meaningful subpopulation in our college students. An indirect interpretation of the groups would be that the groups emerged to account for factor model misspecification, nonnormality, or nonlinearity (Bauer & Curran, 2004). We consider either type of latent class important to uncover. Thus, we recommend that future research investigate further which interpretation is more appropriate for these classes.

Bauer and Curran (2004) provide explicit recommendations about how to determine whether latent classes should be interpreted directly or indirectly. For example, Bauer and

Curran recommend determining the optimal number of classes using a saturated measurement model, that is by definition not misspecified. Since latent classes may emerge spuriously because the factor model was misspecified. In practice, this method may be problematic because unrestricted mixture models have unpredictable likelihood functions, and that model comparison indices penalize for more complex models, so meaningful model comparisons between the models with unrestricted and restricted measurement components may be complicated.

Further, Bauer and Curran caution that latent classes may be appearing because of nonnormality in the data. Bauar and Curran recommend comparing the distributions of the observed scale scores and factor scores to determine how well the model is reproducing the distribution. Also, it is important to consider possible sources of nonnormality or nonlinearity and how those may be contributing to class results. For example, nonnormality could be occurring because the distribution for the latent trait is truly skewed in the population. Alternatively, the observed distribution may be skewed because of a sampling bias that resulted in a ceiling or floor effect.

The finding that items 2 and 6 were more different across classes may deserve further consideration. These two items are the only two items on the MAF scale which use the strong phrasing “afraid of failure” instead of language of “dislike” or “uneasiness.” It may be that this stronger language is interpreted differently by students. Previous research has shown that using fear words in items can result in those items looking more similar to each other than to other items (Miller, 2005; Young, Finney & Lau, 2005). One substantive interpretation of the classes could be that some student react to the items that reference fear differently than other items, while other students do not.

Regardless of the source of the multiple classes, it is clear from this study that something a simple one-factor one-class model may not be representing the data adequately. If the study had found that the one-factor one-class model indices were as good as the two-class model, we would have found strong support for using the one-factor model of the MAF. However, given that we found evidence of multiple classes, more psychometric research on the MAF scores is needed.

Limitations of the study

The current study was limited in that estimation issues prevented us from obtaining stable solutions for more than two class solutions as hoped. Also, although we looked at a set of external variables, they were limited to demographic and motivation variables. Ideally, we would have liked to examine more comprehensive set of external variables. Finally, our study only examined one sample of students. We would have liked to have additional sample available to see if the class solution replicated.

Future Research

There are several avenues for future research in this area. First, because of the risk of identifying spurious classes, the model tested in this study should be re-examined on multiple other samples. Second, other external variables not considered in this study should be examined in future research. Third, the current study examined only two different solutions, when many other models are plausible. Future research should explore alternative models and engage in a more rigorous model comparison process.

Specifically, it would be useful to formally test that measurement invariance of the two classes by comparing the fit of nested models in which different factor parameters are constrained versus free. In this study we allowed all parameters to vary, except for those we had to fix for model identification, and we simply described the differences. Future studies could determine whether the differences we found are statistically meaningful through tests of nested models.

Other models to explore might be to examine a two-factor model of the MAF scale (Young, Finney & Lau, 2005). Our results indicate that there may be two sets of items that differ across classes to varying degrees. It is possible that the latent classes emerged because the factor model was misspecified (Bauer & Curran, 2004). Thus examining the class solution when a two-factor model of the MAF scale is posed may be interesting.

Conclusions

The current study identified the need for further research on the MAF scale for use in a college student population. Latent classes emerged indicating that a program of research is necessary to better understand the causes of the latent groups. Subsequently, the implications for MF theory and use of the MAF scale can be explored. Factor mixture modeling is a powerful but complex technique which can be used to explore the internal structure of a scale. Researchers are cautioned to be aware of the multiple potential reasons that subpopulations maybe found and to address those before using FMM results in practice.

References

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review, 64*, 359-372.

Atkinson, J. W., & Litwin, G. H. (1960). Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. *Journal of Abnormal & Social Psychology, 60*, 52-63.

Ceranski, D. S., Teevan, R., & Kalle, R. J. (1979). A comparison of three measures of the motive to avoid failure: Hostile press, test anxiety, and resultant achievement motivation. *Motivation and Emotion, 3*, 395-404.

Conroy, D. E., & Elliot, A. J. (2004). Fear of failure and achievement goals in sport: Addressing the issue of the chicken and the egg. *Anxiety Stress and Coping, 17*(3), 271-285.

Covington, M. V. (1984). The self-worth theory of achievement motivation: Findings and implications. *Elementary School Journal*, *85*, 5-20.

Diseth, Å, & Martinsen, O. (2003). Approaches to learning, cognitive style, and motives as predictors of academic achievement. *Educational Psychology, 23*, 195-207.

Dweck, C. S., Chiu, C., & Hong, Y. (1995). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry, 6*(4), 267-285.

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., & McGregor, H.A. (2001). A 2x2 achievement goal framework. Journal of Personality and Social Psychology,80, 501-519.

Elliot, A. J., Sheldon, K. M., & Church, M. A. (1997). Avoidance personal goals and subjective well-being. *Personality and Social Psychology Bulletin*, *23*, 915-927.

Fairchild, A. J., Horst, S. J., Finney, S. J., & Barron, K. E. (2005). Evaluating existing and new validity evidence for the academic motivation scale. *Contemporary Educational Psychology, 30*(3), 331-358.

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

Gjesme, T. (1982). Amount of manifested test anxiety in the heterogeneous classroom. *Journal of Psychology: Interdisciplinary and Applied*, *110*, 171-189.

Harackiewicz, J. M., Barron, K.E., Carter, S.M., Lehto, 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.

Hagtvet, K. A., & Benson, J. (1997). The motive to avoid failure and test anxiety responses: Empirical support for integration of two research traditions. *Anxiety, Stress & Coping: An International Journal*, *10*, 35-57.

Heckhausen, H., & Strang, H. (1988). Efficiency under record performance demands: Exertion control--an individual difference variable? *Journal of Personality and Social Psychology, 55*, 489-498.

Miller, B. J. (2005, April). *Examining the avoidance subscales of the Achievement Goal Questionnaire*. Poster presented at the annual meeting of the American Educational

Research Association, Montréal, Canada.

Minnaert, A. E. (2003). The moderator effect of test anxiety in the relationship between test expectancy and the retention of prose. *Psychological Reports, 93*(3), 961-971.

Minnaert, A. E. (1999). Individual differences in text comprehension as a function of test anxiety and prior knowledge. *Psychological Reports, 84*(1), 167-177.

Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus user's guide* (4th ed.). Los Angeles, CA: Muthén & Muthén.

Nygárd, R., & Gjesme, T. (1973). Assessment of achievement motives: Comments and suggestions. *Scandinavian Journal of Educational Research, 17,* 39-46.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461-464.

Sundre, D. L. & Moore, D. L. (2002). The Student Opinion Scale: A measure of examinee motivation. *Assessment Update, 14* (1), 8-9.

Raynor, J. O. (1970). Relationships between achievement-related motives, future orientation, and academic performance. *Journal of Personality and Social Psychology*, 15, 28-33.

Young, H., Finney, S. J., & Lau, A. R. (2005). *Factor structure, measurement invariance, and latent mean analysis of motive to avoid failure between lower- and higher-SES students.* Unpublished manuscript, James Madison University.