Journal of Leisure Research
2003, Vol. 35, No. 2, pp. 129-151
Copyright 2003
National Recreation and Park Association

Articles

Development of the Free Time Motivation Scale for Adolescents

Cheryl K. Baldwin and Linda L. Caldwell
School of Hotel, Restaurant and Recreation Management
The Pennsylvania State University

A new self- report measure of adolescent free time motivation (FTMS-A) based in self-determination theory (Ryan & Deci, 2000) has been developed. The scale measures five forms of motivation (amotivation, external, introjected, identified, and intrinsic motivation) and is appropriate for use with young adolescents (ages 12-15). Using confirmatory factor analysis, examination of each of the motivation subscales indicated minimally acceptable levels of fit. The test of the overall model without modification was also minimally acceptable. The deletion of two items improved the fit and provides preliminary evidence of the validity of the FTMS-A, however, future replication of this finding is needed.

KEYWORDS: *Adolescence, free-time, motivation*

Why do youth do what they do in their free time? If you ask them, typical answers are "I have to," "I want to," or "There is nothing else to do." These simple answers belie the complexity of motivation. Understanding and measuring motivation in terms of adult behavior is difficult enough, but during adolescence the developmental processes through which youth evolve compound this complexity. The purpose of the current study was to advance the measurement of motivation in the free time domain by testing the factorial validity of a multidimensional scale grounded in Deci and Ryan's self-determination theory (1985, Ryan & Deci, 2000). In this study, we focus on measuring early adolescents' motivation for free time activity.

Self-determination theory (Deci & Ryan, 1985, Ryan & Deci, 2000) is a useful framework for understanding varying levels of motivation for activity engagement. It addresses conditions of intrinsic and extrinsic motivation and has been used to study motivated behavior in the educational (Ryan & Connell, 1989), sport (Frederick & Ryan, 1995; Vallerand & Fortier, 1998), and, to a lesser extent, the free time or leisure domains (Pelletier, Vallerand,

Address correspondence to Cheryl K. Baldwin, 201 Mateer Building, University Park, PA 16802. Phone: (814) 863-0015. Fax: (814) 863-4257. Email: ckb10@psu.edu.

Author Note: The authors gratefully acknowledge Tanya Boone, Susanne Dubrouillet, and John Persing for their many contributions to this project. This paper is based upon work supported by the U.S. Department of Health and Human Services under Award No. 1R21 DA13193-01 awarded to L. Caldwell. This project is totally financed by federal funds in the amount of $420,900.

Green-Demers, Blais, & Briere, 1996). Much of Deci and Ryan's original work on self-determination theory used a laboratory setting and experimental framework. Since then, a number of researchers have adopted self-determination theory as a framework for studying motivated behavior in natural settings and have used a modified version of Ryan and Connell's (1989) children's (grades 3rd-6th) academic motivation self-report to study motivated behavior.

We present in this article a scale, the Free Time Motivation Scale for Adolescents (FTMS-A), which follows the framework for measurement established by Ryan and Connell (1989) and Pelletier et al. (1996). The FTMS-A differs from Ryan and Connell (1989) in that it applies to the free time rather than the educational domain. It also differs from Pelletier et al. in that the measure applies to a younger population of middle school age students (6^{th}-8^{th} grade) rather than senior high students (9^{th}-12^{th} grade).

The need to develop this scale reflects several advancements in the understanding and measurement of motivated behavior from the self-determination framework. First, the measurement of intrinsic and extrinsic motivation has moved beyond a dichotomous interpretation (Rigby, Deci, Patrick & Ryan, 1992). This dichotomy has been especially prevalent in research in sport (Vallerand & Fortier, 1998). Following Deci and Ryan (1985), the assumption was that sport activities were imbued with intrinsic qualities (Frederick & Ryan, 1995). This resulted in studies within sport that measured levels of intrinsic and extrinsic motivation, but tended to result in the loss of emphasis of the differentiated framework between intrinsic and extrinsic motivation (Vallerand & Fortier).

Understanding and measuring levels of motivation for free time activity among youth is particularly important because of the developmental demands associated with autonomy development and the processes of individuation and differentiation that characterize the period of adolescence. The uncomplicated reasons of "I have to, want to, or nothing else to do" evolve into more complex and contextual reasons of "I feel obligated," or "If I do this, then I might get [some reward] in the future." Free time is a naturally productive context from which to understand motivational processes because the inherent nature of the context allows for a full range of motivations to occur (i.e., the most intrinsically and extrinsically motivated behaviors are possible). Parsing these types or levels of motivation into a useful measure is an important step in research efforts undertaken to explore adolescent use of free time.

Conceptual Framework

Self-Determination Theory

Self-determination theory (Deci & Ryan, 1985) is a model of human motivation rooted in the innate needs of competence, relatedness, and autonomy and it is an overriding framework for intrinsic and extrinsic motivation. Fulfillment of these needs is a prerequisite for optimal human functioning. Self-determination theory (SDT) addresses the natural human tendency to actively engage in the world and is a framework for investigating the social or environmental factors that enhance or forestall innate needs. The motivational processes that underlie self-determination are elaborated in two subtheories of SDT (Deci & Ryan, 1985). Cognitive evaluation theory (CET) is a subtheory of SDT that explains variability in intrinsic motivation. Organismic integration theory (OIT) addresses the self-regulatory process associated with varying forms of extrinsic motivation. Intrinsic motivation refers to the natural and inherent tendency to seek out novelty, challenge, pursue interests, and engage in activity as an end in itself. In contrast, extrinsic motivation refers to engagement in an activity as a means to an end or due to some external compulsion.

Cognitive evaluation theory conceptualizes intrinsic motivation as the prototypic form of self-determined, motivated, intrinsically regulated behavior (Ryan & Deci, 2000). Since intrinsic motivation is an inherent tendency produced by innate needs, CET addresses the social and environmental conditions that support or hinder the expression of intrinsically motivated behavior. Organismic integration theory addresses the social and environmental factors related to amotivation and the processes of internalizing and integrating different forms of extrinsically motivated behavior. These theories combine in an overall understanding of self-determined behavior characterized by intrinsic and extrinsic motivation. Forms of self-determined behavior vary in regulatory style and perceived locus of causality.

A Multidimensional Conceptualization of Intrinsic and Extrinsic Motivation

Intrinsic and extrinsic motivations are conceptualized as opposites on a continuum of self-determined behavior (Figure 1). Within this continuum, six forms of motivation have been identified (Deci & Ryan, 1985). At one end, intrinsic motivation represents behavior characterized by the pursuit of an activity that one finds interesting and is engaged in for the inherent satisfaction derived from participation. At the other end of the continuum

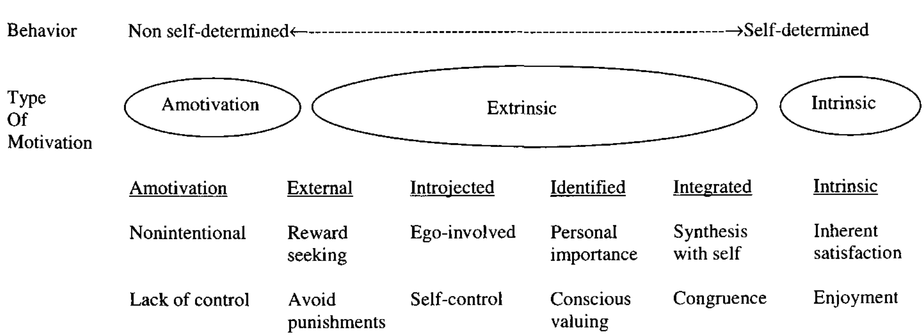

Figure 1. Continuum of types of motivation (adapted from Ryan & Deci, 2000)

is amotivation, which describes behavior that is nonintentional and nonregulated. Between the two ends are four forms of extrinsically motivated behavior: external, introjected, identified, and integrated. Although the latter two forms of extrinsic motivation, identified and integrated, are characterized by external forces, there are elements of intrinsicness and self-direction related to these forms of motivation.

The forms of extrinsically motivated behavior vary in relative autonomy and regulatory style. The most extreme form of extrinsic motivation, external regulation, refers to doing an activity to satisfy external demands. One perceives the cause of action as external to the self and action is motivated by receiving rewards or avoiding punishments. An adolescent who is externally regulated may be participating on a basketball team only because her father demands that she do so.

Introjected regulation is motivation based on a *self-controlled, ego-involved* form of behavior that is typically driven by a perception of what others might think. Introjected motivation represents actions that are carried out based on contingencies, for example, one acts to avoid guilt or anxiety. Similarly, doing something to maintain one's status or feelings of self-worth is also considered introjected regulation as such action is not autonomously based, but rather, externally focused. Although somewhat simplistic, adolescents motivated by peer pressure are exhibiting introjected regulation. A boy who learns to play the guitar and joins a rock band mainly to be perceived as "cool" by his friends is an example of this form of motivation.

Identified regulation refers to a regulatory style that is somewhat internal and is associated with behavior that is motivated by the feeling that engaging in an activity is the result of valuing the activity. One views action as personally important, therefore engagement is relatively autonomous. This type of motivation is goal oriented and there is a purpose associated with engagement. For example, an adolescent who joins the debate team may find it enjoyable and intrinsically rewarding, but if the main reason for participation is to prepare him to do well in college, he is "identified" in his motivation for this activity.

Integrated regulation refers to motivation that is based in action congruent with one's own beliefs. One has evaluated the regulatory process and assimilated it to be in correspondence with one's other values and needs. It is the most autonomous form of extrinsically motivated behavior and produces benefits associated with intrinsic motivation, such as interest and enjoyment. It is difficult to provide a youth-based example of this form of motivation because adolescents are not yet cognitively and developmentally capable of this form of motivation (Vallerand, 1997).

The conceptual importance of teasing out these various forms of motivation in free time remains to be seen. Theoretically, these forms of motivation would predict, among other things, enjoyment of activity, healthy behaviors, level of constraint, and ability and desire to overcome challenges associated with participation in leisure—all aspects of healthy adolescent development. This methodological paper details our attempt to measure these various forms of free time motivation. It is important to note that our scale is context and not activity specific. That is, the entire free time context is the frame for understanding motivation, so we present a general motivation scale. Most other nonacademic efforts have been activity specific and thus are useful only to a specific activity domain (e.g., basketball).

Focusing the scale on free time as a context rather than on specific free time activities has both benefits and consequences, not unlike other measures that could be global, contextual, or domain specific (e.g., coping, health, quality of life, satisfaction, and so on. See Goossens, 2001, for a discussion on identity as a global versus domain specific measure). We choose to measure motivation at the contextual level because the other variables (outcome and mediator) in the larger study from which this paper derived were also measured at the free time context level (e.g., boredom in leisure, decision making and planning in leisure). As well, the leisure education program that was the focus of the evaluation for the larger study was conceptualized at the "free time" level. The intention of the intervention was to assist adolescents in being responsible for managing the whole of their free time. Thus, conceptual congruence among levels of measurement and intervention focus was achieved in the entire study. Our decision to measure general free time motivation among adolescents comes with some conceptual and methodological challenges, however. These, and the details of our analysis, are presented in the following sections.

Method

Developing the FTMS-A was an essential part of a larger effort designed to evaluate the effectiveness of a leisure education program (*TimeWise:* Learning Lifelong Leisure Skills) for reducing or preventing the onset of substance abuse among middle school youth (Caldwell, Smith, Swisher, & Vicary, in progress). To evaluate *TimeWise,* a three-year, quasi-experimental design is being employed, and both outcome and process data are being collected. Nine school districts in central Pennsylvania are participating in the evaluation, four of which receive the *TimeWise* program and five serve as comparison schools. The ultimate outcome of *TimeWise* is to prevent or mitigate the onset of problem behavior (e.g., substance abuse). A number of leisure-related mediators were posited to influence the ability of youth to partake in healthy versus unhealthy leisure behaviors (e.g., boredom in leisure, motivation, decision making and planning skills, persistence, ability to overcome constraints, interest development). The evaluation of *TimeWise* is beginning its third year, and three waves of data have been collected. The data used in this study come from the baseline data collected in fall 2000.

Questionnaire Construction

The Free Time Motivation Scale for Adolescents (FTMS-A) is a self-report measure for adolescents based in self-determination theory that generates a score for five types of motivation (amotivation, external, introjected, identified, and intrinsic motivation). No scale was developed for the integrated form of extrinsic motivation, as the construct is not expected to be evident until late adolescence or early adulthood (Vallerand, 1997). The FTMS-A was adapted from the Pelletier et al. (1996) leisure scale for high school students, the Self Regulation Questionnaire developed by Ryan and Connell for elementary students for the education and social domains, the Academic Motivation Scale (Vallerand et al., 1992), and measures of motivation in the sport domain (Chatzisarantis, Biddle, & Meek, 1997; Goudas, Biddle, & Fox, 1994). The FTMS-A differs from prior scales in that it addresses the general free time context and was developed for young to middle adolescents (e.g., ages 12 to 16). The modifications for this context and population resulted in a self-report measure in the tradition of those previously developed, but the instructions, sentence, and item content were notably modified.

The FTMS-A scale was developed in the following manner. An item pool based on the work of Deci and Ryan (1985), Pelletier et al. (1996), Ryan & Connell (1989), and Vallerand et al. (1992) was generated. Of concern to the research team was how understandable, valid, and readable the items of this scale, and others in the study, were. Therefore, a series of cognitive interviews (Willis, 1994) were conducted with a convenience sample of (eight) youth, aged 12 to 16. Two members of the research team asked each adolescent to read the items in the FTMS-A and respond using the response scale provided (the Likert type response scale to be used in the questionnaire). This process mirrored what the youth would be asked to do in the baseline data collection protocol. After the youth completed the questionnaire, the researchers asked the youth questions about each item. These questions encouraged the youth to describe problems they had understanding the items and ideas they had about wording as well as face validity (that is, would the question make sense to young adolescents?). Typical questions were "Which items on the first page seemed awkward or do not make sense to you? Were there any items that you did not understand? Do you have any suggestions as to how we could make the wording more understandable?" During the discussion the researchers probed and discussed the concerns of the youth about each question. Each youth's response was recorded, and after all eight youth went through this process, the entire research team examined each item vis-à-vis the feedback from the youth. From this process, a revised item pool was generated.

It was during this cognitive interview phase of item development that we wrestled with the issue of how to measure a contextual construct such as motivation in free time, rather than a more activity specific construct such as motivation for basketball as representative of sport. The first few youth who participated in the cognitive interviews indicated that it was difficult for them to respond to the items because "their motivation depended on what they were thinking about in terms of their own activities." After much discussion and consultation with measurement and adolescent development experts, as well as the youth who participated in the cognitive interviews, we handled this dilemma by deciding to provide very clear directions before the self-administered questionnaires were distributed in class. Questionnaire administrators were trained to begin their instructions by asking youth to think about *everything* they did in their free time. Examples such as "TV, sports, hanging out with friends" were written on the board. Students were then asked to respond to the items on the questionnaire in terms of "how they generally feel" about everything they do in their free time, and not to just think of how they feel about one particular activity. While this solution was not perfect, responses from the youth in the cognitive interviews and the structure of the data suggest that we were effective in encouraging students to think broadly about their free time in general.

The FTMS-A scale items are listed by motivation category in Table 1. The scale items represent reasons that complete the sentence "I do what I do in free time because. . ." Respondents rated their level of agreement with the reason using a 5-point Likert scale anchored from strongly disagree to strongly agree. The advantage of this format is that focusing on why an individual does something matches the underlying action orientation implied by the concept of motivation (Ryan & Connell, 1989; Vallerand, 1997; Vallerand & Fortier, 1998). Scales that assess intrinsic motivation by asking respondents to indicate their level of interest or enjoyment confound the antecedent conditions of motivation and the expression of intrinsic motivation (Vallerand, 1997).

Participants

Baseline data were collected in September and October, 2000 from 634 grade seven students. Self-report questionnaires were administered in the classroom by a team of trained university students. Of the 634 students at baseline, 315 were female (49.7%) and 95% percent of respondents were Euro-American. We received parental permission and collected data from between 51% and 88% of all grade seven students in each of nine schools (the average was 63%). Ninety-nine percent of all students with parent permission completed the self-report questionnaire.

Strategy for Data Analysis

Initial descriptive and preliminary analyses were first employed to assess the internal reliability of the scales, to diagnose and deal with missing data, and assess assumptions of normality. After that, both correlational and confirmatory factor analyses were conducted to assess the internal validity of the multidimensional scale. Given that the factor structure for the FTMS-A was expected to conform to the well-defined theoretical structure of SDT, confirmatory factor analysis (CFA) was employed to assess the measurement model for each factor prior to testing the first order factorial structure of the model. While the items represent modifications from previous scales, the

TABLE 1
Free Time Motivation Scale for Adolescents

In this survey, we are asking you to think about your free time. Free time means things that you do outside of school. These can include after-school activities like sports or clubs, and activities like 4H, music, spending time with friends, reading, and watching TV.

Directions: Circle the answer that best reflects WHY you do what you do in your free time

I DO WHAT I DO IN MY FREE TIME BECAUSE. . . .

Amotivation (AMT)
- AMT1 I don't know why I do my free time activities, and I don't really care.
- AMT2 I don't know, nothing much interests me.
- AMT3 I don't know, I have never really thought about it.
- AMT4 I don't know but it doesn't matter because I don't do much of anything.

External Motivation (EXT)
- EXT1 I would get in trouble if I don't.
- EXT2 I am supposed to.
- EXT3 That is the rule in my house.
- EXT4 So others won't get mad at me.
- EXT5 My parents expect me to.

Introjected Motivation (IJ)
- IJ1 I want people to think I am good at what I do.
- IJ2 I will feel badly about myself if I don't.
- IJ3 I want to impress my friends.
- IJ4 I want people to like me.
- IJ5 I want to earn rewards, medals, trophies, or certificates.

Identified Motivation (ID)
- ID1 I want to understand how things work.
- ID2 What I do is important to me.
- ID3 I develop skills that I can use later in life.
- ID4 The activities help me develop into the person I want to become.

Intrinsic Motivation (INT)
- INT1 I want to have fun.
- INT2 I enjoy what I do.
- INT3 I like what I do.
- INT4[a] Sense of freedom.
- INT5 I want to.

[a]item eliminated from scale because of low reliability

theoretical and empirical support for the multidimensional nature of intrinsic and extrinsic motivation suggested that CFA was the most appropriate means for assessing the FTMS-A scale.

Preliminary Analyses

Reliability Analyses

To assess the reliability of each of the subscales Cronbach's alpha was computed for the amotivation, external, introjected, identified, and intrinsic subscales. The four-item amotivation and identified subscales demonstrated satisfactory reliability with coefficient alpha scores of .70 and .67, respectively. The 5-item external motivation subscale was the most reliable with an alpha level of .79. Introjected motivation was also a 5-item subscale and its associated alpha level of .69 also met an acceptable standard. The 5-item intrinsic motivation subscale initially produced a low reliability coefficient (alpha = .68). Reliability analysis indicated that reliability would be improved with the deletion of the item "sense of freedom." Whereas the other intrinsic motivation items conveyed enjoyment and desire, the adolescents may have interpreted freedom as lack of restrictions rather than choice. Therefore, this item was deleted from the scale and resulted in an improved reliability coefficient of .72.

Missing Data Analysis

The missing value analysis procedure for SPSS 10.1 was used to identify the percentage of missing data for the motivation items. Of the 634 students who participated in the study, six students failed to complete more than half of the motivation items (there were 22 items in total). These six students were subsequently removed from the sample for these analyses. An additional 100 students who had missing data on at least one of the motivation items and two students missed three of the 22 items. All of the remaining students with missing data missed only one or two items. Considering missing data by variable, there was no motivation variable item with more than 3% percent missing data. Since the missing data were well scattered across variables and cases there was support for concluding that missingness was random. Although there were enough missing data spread across cases and variables to be concerned with listwise drop in sample size, there is no clear best course of action for managing missing data (Tabachnick & Fiddell, 2001). Therefore, we replaced missing data with the expectation maximization procedure available in SPSS 10.1.

Univariate and Multivariate Normality

One of the key assumptions of CFA is that the data follow a multivariate normal distribution. Therefore, item distributions were assessed for normality. Univariate normality is a prerequisite, though not sufficient, condition for multivariate normality (West, Finch, & Curran, 1995). Statistical tests of skewness, kurtosis, and univariate normality were employed in combination with visual screening of the item histograms and stem-and-leaf diagrams. The amotivation and external items were positively skewed (2.18 to 14.30) and the identified and intrinsic items were negatively skewed (−7.08 to −26.26). The introjected scale was comprised of two positively skewed (1.93 & 3.00) and three negatively skewed (−5.95 to −8.16) items. The distributions for all but three of the twenty-two items displayed a non-normal kurtosis. Results of the Lilliefors test confirmed that the data did not display a univariate normal distribution.

Although most social science data are nonnormal (Micceri, 1989), most researchers have ignored the problem of nonnormality (Byrne, 2001). The problem with ignoring nonnormal data distributions in CFA is that the chi-square statistic becomes excessively large and models with good fit are rejected greater than 5% of the time. In addition, standard errors are underestimated biasing the parameter estimates and resulting in too many significant results (West et al., 1995, p. 59). This issue is discussed further in the section describing the results of the confirmatory factor analysis.

Results

Correlational Analyses

One approach to the validation of the multidimensional motivational framework is to assess the intercorrelations between the motivational subscales. The forms of motivation, amotivation, external, introjected, identified, and intrinsic, are hypothesized to correlate in a simplex-like structure consistent with the theoretical self-determination continuum (Ryan & Connell, 1989). To determine a simplex structure, correlations are visually inspected to assess whether scores for the subscales of motivation that are theoretically more similar (i.e., closer together on the motivation continuum) correlate more strongly and positively than those that are more distant along the continuum.

As shown in Table 2, FTMS-A subscale correlations demonstrate the hypothesized simplex-like structure. In addition, amotivation and intrinsic motivation display an expected negative and significant correlation. These results compare favorably to those of Ryan and Connell (1989) and support the construct validity of the FTMS-A.

Confirmatory Factor Analysis

To evaluate the factor structure of the FTMS-A, the following steps were taken. First, the single factor structure for each subscale was assessed. The purpose of this step is to confirm the measurement qualities of each subscale. It is expected that each subscale item will have a nonzero loading on the subscale. An insufficient measurement model at the subscale level would

TABLE 2
Simplex-like Structure of Sub-scale Item Correlations

	AMT	EXT	IJ	ID	INT
Amotivation	—	.266**	.053	−.300**	−.359**
External		—	.549**	.141**	−.192**
Introjected			—	.397**	.062
Identified				—	.480**
Intrinsic					—

**significant at .01 level

indicate that the subscale is not adequately measuring the latent factor and that the subscale is not a unique facet of the multidimensional 5-factor structure. The second step in the confirmatory analysis was to test the measurement model and factor structure of the five-factor FTMS-A model. It was expected that the five factors would correlate in a simplex-like manner and each subscale item would retain a nonzero loading on the factor it was designed to measure and a zero loading on all other factors.

Before these steps were taken we had to address the nonnormality of the data. The nonnormal character of the data warranted consideration as it violated one of the key assumptions of CFA. One of the most straightforward means of dealing with nonnormal data is to use a square root or log transformation so that the data more closely approximate the normal distribution. When these transformations were employed to our data, the item distributions did not markedly improve.

Therefore, our attention turned to the robustness of the confirmatory factor statistics. Maximum likelihood (ML) is one of the most common structural equation modeling estimation procedures. Underlying assumptions of ML are that the data be multivariate normal and from a continuous scale. The robustness of the ML estimate to violations of these assumptions have been investigated and early evidence indicated that, as the data become increasingly nonnormal, the (a) the chi-square statistic value may be inflated, (b) some fit indexes are modestly underestimated (Tucker-Lewis fit index, comparative fit index), and (c) standard errors of the parameter statistics are underestimated (Byrne, 2001; Nevitt & Hancock, 2001; West et al., 1995). More specifically, nonnormality can lead to low standard errors for the estimated parameters suggesting that factor loadings, covariances, and variance estimates may be statistically significant although they may not be in the population (Byrne, 2001).

There are several remedies for compensating for the inflated model statistic and attenuated standard errors associated with nonnormal data. Some remedies, such as rescaling the chi-square statistic and standard errors, are software specific (Tabachnick & Fidell, 2001). Other estimation techniques require sample sizes of 1,000 or more (West et al., 1995). The study sample size and the degree of nonnormality indicated that a corrective strategy was needed. Amos 4.0 software, which was used to analyze the data, provides two applicable techniques, bootstrapping and the Bollen-Stine adjusted chi-square.

Bootstrapping is a computer generated resampling technique. Multiple new samples are created from the research sample to produce a bootstrap (empirical) sampling distribution "which technically operates in the same way as does the sampling distribution generally associated with parametric inferential statistics" (Byrne, 2001, p. 269). The Amos 4.0 software provides an average bootstrap sampling distribution value for the regression coefficient, multiple squared correlation, variance, and standard errors. The rationale for the use of bootstrapping is that, when the assumptions associated with a statistical test are violated, it is better to create an empirical distribution, which is not restricted by the assumption of normality (Zhu, 1997). This allows the ML estimates to be assessed alongside the bootstrap ML standard errors.

Specifically of interest here are the bootstrap estimates of the standard errors for the parameters with a 90% bias corrected confidence interval for the parameter. The confidence interval is associated with a significant test of the respective parameter (regression weight, squared multiple correlation, and variance) and is interpreted in the usual manner. For example, in the case of the regression weight, the null hypothesis of the statistical test is that the regression coefficient is equal to zero. If the confidence interval does not include zero then the hypothesis is rejected. For the full five factor model, we report an associated *p*-value that indicates how small the confidence interval for the parameter would need to be to include zero. For example, a parameter *p*-value of .005 implies that the confidence interval would have to be at 99.5% level before the lower bound value would be zero.

In addition, the Bollen-Stine bootstrap chi-square is reported, which is the adjusted chi square statistic testing the null hypothesis that the specified model is correct. It is interpreted in the same manner as the chi-square for the model without the correction.

Maximum likelihood estimation along with bootstrapping (AMOS 4.0) was employed to test the measurement model for each of the motivation subscales. Since bootstrapping can fail with insufficient sample size we elected not to split the sample for purposes of replication. Each of the motivation subscales was examined for fit prior to testing the overall multidimensional free time motivation model.

The same form of specification was used for each of the single factor subscales. Each motivation latent factor (i.e., the subscale construct) has a directional influence on the respective measured indicator variables. Bootstrapping requires that one factor loading path per factor be constrained to some nonzero value. The regression weight for the path from the latent factor to the first item was fixed to one. The error variables associated with each indicator variable represent the unexplained unique error variance, reflecting measurement and random error; the regression weights for these paths were also fixed to one. The same specification format was used in the full five factor model analysis.

Support for the hypothesized model is established by a good fit between the observed and predicted covariance matrices. Fit is assessed statistically with the chi-square goodness of fit test and through a number of fit indexes. Following Byrne (2001) and to represent a range of assessment strategies, we report the overall model chi-square value and the values for a number of the established fit indices. We briefly review the indices reported.

Chi-Square and chi-square/degrees of freedom ratio. The chi-square goodness of fit statistic is a test of discrepancy between the predicted and observed models, a nonsignificant result is a finding in support of the model fit. It is well established that the chi-square goodness of fit test is sensitive to sample size (Tabachnick & Fidell, 2001). As sample size increases it is not uncommon to reach statistical significance even with small differences between the observed and predicted models. Given this fact, numerous researchers suggested using the ratio of the chi-square value divided by the model degrees of freedom (minimum discrepancy, CMIN) as an additional indicator of fit. The interpretation of an acceptable value of the CMIN ranges from 5 to 1 to 2 to 1 (Arbuckle & Wothke, 1999).

Comparative Fit Index (CFI). The comparative fit index (Bentler, 1990) is one of the most commonly reported fit indices. This index uses a baseline model for comparison purposes, meaning that the fit is examined in regard to an independence model of fit, which is the standard of no fit at all. The measure varies from zero to one with one indicating a perfect fit and the general rule of thumb for minimum acceptable fit is .90. Recently, Hu and Bentler (1998) argued that .95 be replaced as the general standard for minimum fit.

Parsimony Comparative Fit Index (PCFI). The PCFI measure of fit addresses the complexity of the model. It tends to be lower than the average fit index with values of approximately .50 serving as a baseline value (Byrne, 2001). Parsimony indexes take into consideration the simplicity/complexity of the model. Inclusion of parsimony indices is recommended because freeing up more parameters in the model can increase the value of the goodness-of-fit indices. Therefore, parsimony based measures attempt to build in consideration of model complexity relative to parameters estimated and degrees of freedom.

Root Mean Square Error of Approximation (RMSEA). The root mean square error of approximation is a measure based on population discrepancy. It estimates the lack of fit using a perfect or saturated model for comparison. A saturated model is a test that compares the observed fit to a model that fits the data perfectly. The rule of thumb for RMSEA is that values of .05 or less indicate a close fit, .06-.08 a reasonable fit, and values of .10 the upper limit of acceptable fit (Arbuckle, 1999; Byrne, 2001). Also reported is the 90% confidence interval for this statistic (LO 90 – HI 90). Narrower confidence intervals are indicative of good precision of the RMSEA value. Finally, PLCOSE represents the *p*-value for testing the null hypothesis that the population value of RMSEA is less than or equal to .05. Therefore, a nonsignificant PCLOSE value indicates support for model fit.

Single Factor Structure—Motivation Variables

Table 3 displays the overall chi-square and indices of fit for each of the motivation factors. The standardized regression coefficients, the squared multiple correlations, and the variances for the latent factors (i.e., subscales) and error terms are displayed in Table 4. All were significant at $p \leq .01$. The chi-square tests were significant for each of the motivation subscales except identification. Because of the moderate sample size and nonnormal data these significant chi-square values were expected. The CMIN were acceptable (<5 to 1) with the exception of the values for the amotivation (7.14). The CFI values for each of the motivation factors were all well above the minimum acceptable minimum standard of .90. Likewise, the RMSEA values were acceptable (<.10) for all with exception of amotivation (.10). This conclusion was further supported by the PCLOSE values. The fit indexes related to parsimony are generally low (<.50). The PCFI values for the external (.49) and introjected (.49) factors were at marginally acceptable (.50) values.

Also illustrated in Table 3, the Bollen-Stine bootstrap (BS_{boot}) of model fit was nonsignificant for the external, identification, and intrinsic factors indicating that in accord with the other fit measures, these models provide adequate to good fit. The Bollen-Stine bootstrap value was significant (<.05) for the amotivation and introjected factors. For the amotivation model, this led to additional concern for the acceptability of the model. With the introjection model, it raised some concern as it contrasted with acceptable indicators of model fit.

In cases of a lack of fit, the differences between the predicted and observed residual covariance scores provide insight into the nature of the lack of fit. While most of the fit indices for the amotivation model suggested a

TABLE 3
Fit Indices for Motivation Subscales

	AMT	EXT	IJ	ID	INT
χ^2	14.28	14.40	18.04	1.28	9.00
χ^2 *p*-value	.00	.01	.00	.53	.01
χ^2/df (CMIN)	7.14	2.88	3.61	.64	4.50
CFI	.97	.99	.97	1.00	.99
PCFI	.32	.49	.49	.33	.33
RMSEA	.10	.06	.06	.00	.08
LO	.06	.02	.03	.00	.03
HI	.15	.09	.10	.07	.13
PCLOSE	.04	.35	.19	.85	.16
BS_{Boot}	.01	.07	.02	.53	.15

TABLE 4
Mean, Standard Deviation, Standardized Regression Weight, Squared Multiple Correlation and Variance Scores for the Motivation Subscales

Variable	*M*	*SD*	β[a]	R^2	σ^2[b]
AMT1	2.13	1.23	0.54	0.29	0.87
AMT2	1.74	0.99	0.69	0.48	0.50
AMT3	2.25	1.12	0.59	0.35	0.81
AMT4	1.71	0.90	0.65	0.42	0.47
AMT	1.96	0.75	—	—	0.35
EXT1	2.22	1.17	0.50	0.25	1.03
EXT2	2.19	1.14	0.69	0.48	0.66
EXT3	2.21	1.18	0.77	0.59	0.56
EXT4	2.40	1.19	0.63	0.40	0.85
EXT5	2.73	1.31	0.70	0.49	0.87
EXT	2.35	0.88	—	—	0.34
IJ1	3.78	1.10	0.52	0.27	0.87
IJ2	2.56	1.19	0.40	0.16	1.18
IJ3	2.69	1.23	0.66	0.44	0.84
IJ4	3.49	1.23	0.72	0.51	0.76
IJ5	3.74	1.20	0.46	0.21	1.13
IJ	3.25	0.79	—	—	0.33
ID1	3.78	0.91	0.46	0.22	0.64
ID2	4.11	0.90	0.57	0.33	0.54
ID3	3.94	0.85	0.68	0.46	0.39
ID4	3.96	0.91	0.63	0.39	0.50
ID	3.94	0.63	—	—	0.18
INT1	4.67	0.61	0.54	0.29	0.26
INT2	4.41	0.76	0.75	0.56	0.26
INT3	4.35	0.76	0.71	0.50	0.29
INT5	4.30	0.92	0.55	0.30	0.58
INT	4.43	0.56	—	—	0.11

[a] represents the path from latent factor to the item
[b] represents variance for error term

lack of fit, the standardized residual covariances were all below 2.58, the critical value for statistically significant differences between the predicted and observed models (Joreskog & Sorbom, 1988 as cited in Byrne, 2001, p. 89). This indicated that the lack of fit may be an artifact of the statistical test. Modification indices also offer insight into fit, however the parameter change values for the modification values were all quite low. The highest modification index indicated that freely estimating the covariance between the error terms associated with the first and second amotivation error terms (E-AMT1, E-AMT2) would improve the overall chi-square and result in a negative covariance between E-AMT1 and E-AMT2. This is theoretically acceptable as the "don't really care" component of item one suggests a general disaffected state about activity whereas item two implies a lack of interest.

The introjected subscale also had a significant Bollen-Stine bootstrap value, which contrasted with other acceptable fit indices. Analysis of the standardized residuals indicated that none of the values exceeded the 2.58 critical value. This suggested that despite the Bollen-Stine bootstrap value, the observed data fit the predicted model.

In sum 3 of the 5 motivation subscales demonstrated acceptable levels of fit. The fit of the amotivation and introjected models were less clear. The decision to modify these models at this point in the analysis was not clear-cut. Given the mixed evidence and that these models were operating at a low level of degrees of freedom, we proceeded to test the full five-factor model without modifying the amotivation and introjected subscales.

Test of the Five-Factor Motivation Model

Figure 2 displays the results of the confirmatory factor analysis of the first-order, five-factor model, twenty-two item version of the FTMS-A. All parameter estimates were significant at $p \leq .01$ level. The bootstrap standard errors and p-values associated with the 90% bias corrected confidence intervals were also inspected. The largest p-values for the standardized regression weights, variances, and squared multiple correlation were .044, .034, and .044 respectively, meaning that the estimate for the parameters would have to be at the 95.6% and 96.6% levels before the lower bound values would be zero.

The overall chi-square was highly significant (199, N = 628) 561. 72, p = .000 as was the Bollen-Stine bootstrap chi-square (p = .000). The CFI indicated a minimal level of fit of .901. The parsimony based indexes PCFI (.776), and CMIN (2.89) all show adequate fit. The RMSEA values were also acceptable (.054, low = .049, high = .059; p-close = .108).

Review of the standardized residuals and modification indices indicated that introjected items IJ1 and IJ5 were sources of discrepancy. Both items cross-loaded on all of the other subscales (amotivation, external, identified and intrinsic). Given that the CFI was at the minimum acceptable fit and these cross loadings were somewhat problematic theoretically, the decision was made to try and improve the model by dropping these two items.

The analysis was conducted a second time with Items IJ1 and IJ2 removed. The overall chi-square (160, N = 628) = 400.18, p = .000 and Bollen-Stine boostrap chi-square (p = .005) were still significant. However, several of the fit indices were notably improved. The CMIN (2.50) was reduced and the CFI (.928) and PCFI (.781) increased. The RMSEA values were also slightly improved (.049; low = .043, high = .055; p-close = .606). Because this is not a nested model it would be inappropriate to examine the improvement with a chi-square change significance test.

The regression weights, squared multiple correlation, and variances, for the twenty item model, are displayed in Figure 3, all are significant at the

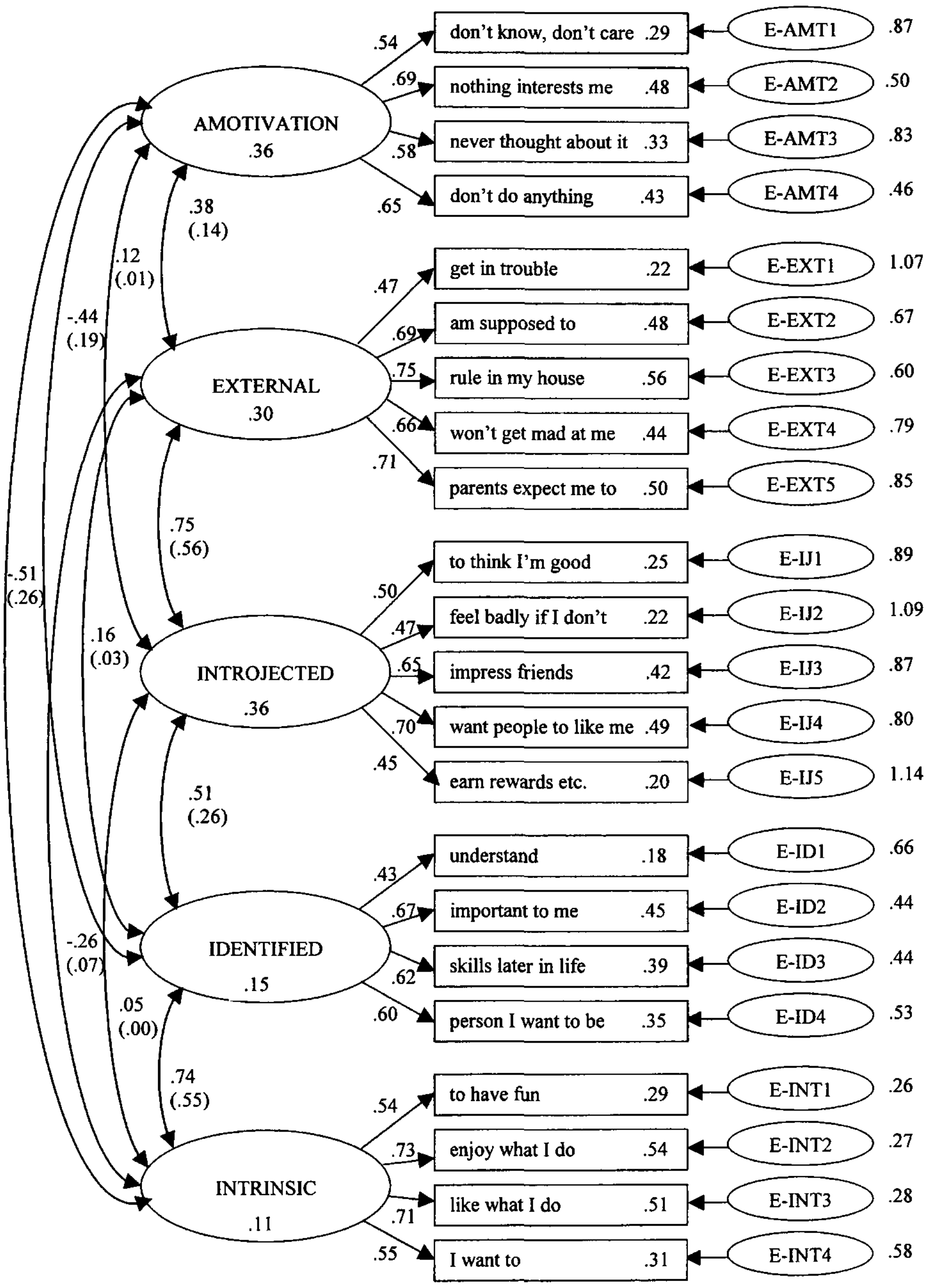

Figure 2. Twenty-two item five factor motivation model

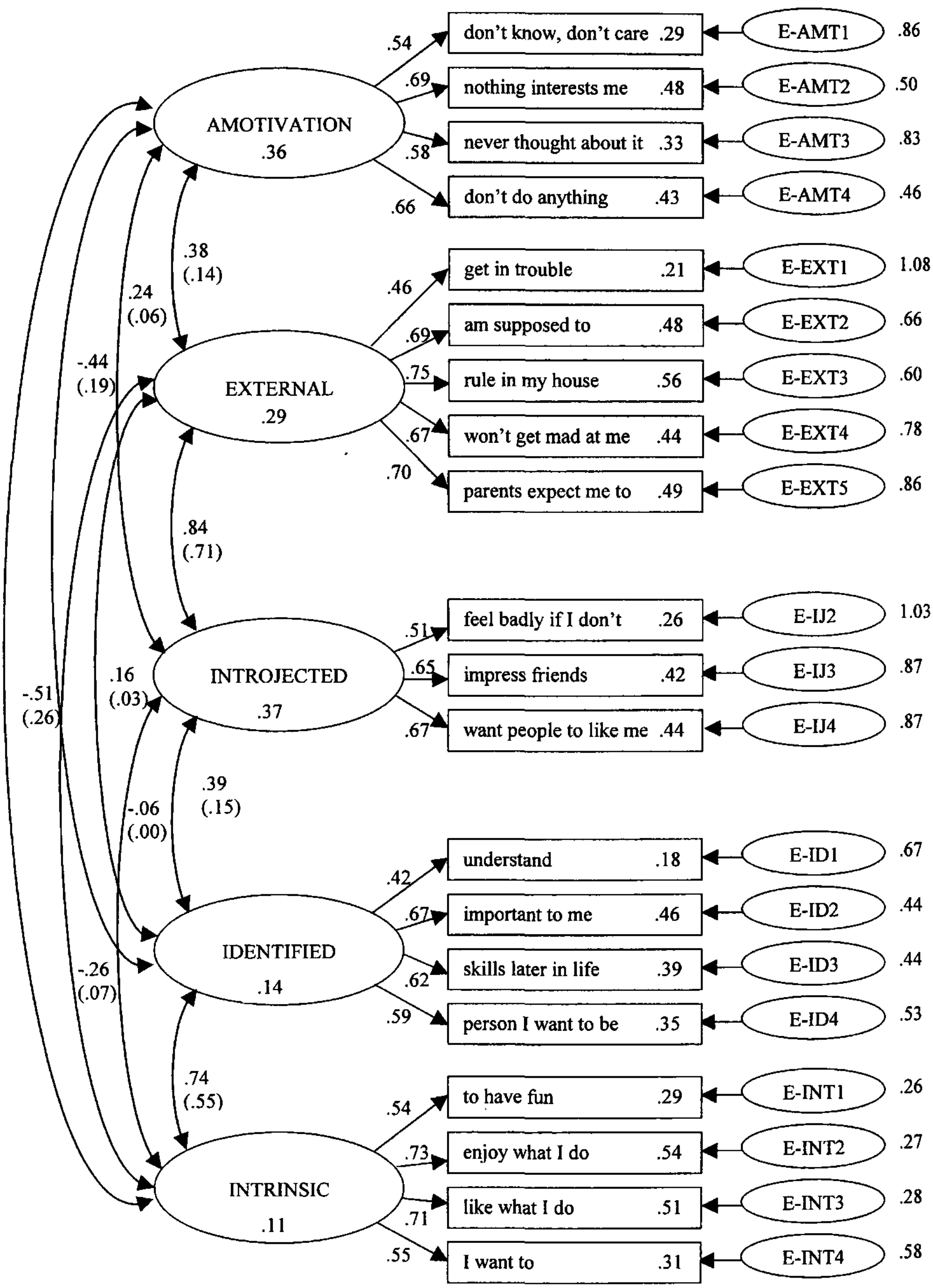

Figure 3. Twenty item five factor motivation model

$p \leq .01$ level. The factor loadings are all adequate and demonstrate loading onto the appropriate latent factor. The values of the squared multiple correlations are also reasonable.

Table 5 provides the composite reliability and variance extracted values for both five factor models. Variance extracted is "the amount of variance captured by an underlying factor in relation to the amount of variance due to measurement error" (Hatcher, 1994, p. 331). Ideally, the variance extracted would be greater than .50 indicating that variance captured by the factor is greater than measurement error. Comparing the twenty and twenty-two item models illustrates that the variance extracted increased for three factors (EXT, IJ, & ID) in the twenty-two item model. However, for the IJ factor there is a variance extracted, reliability trade-off as the reliability of the IJ factor is reduced but variance extracted increases.

Since the correlations between factors are high it is also important to consider whether they are distinct. Following Fornell & Larcker (1981) and Hatcher (1994), the average variance extracted is compared to the square of the correlation (r^2) between the two factors. If the variance extracted for both factors is greater than the r^2 of each, discriminant validity is supported. The r^2 values are identified in brackets underneath the correlation values in Figures 2 and 3. This assessment indicates that EXT-IJ and ID-INT do not demonstrate adequate discrimination.

As a result, we also examined the confidence intervals for the correlations between the factors. Confidence intervals were calculated by multiplying the standard error by two. This value was then added and subtracted from the correlation between the factors. According to Hatcher (1994) if the confidence interval does not include 1.0, discriminant validity is supported. For all factor correlations for both models, neither the upper nor lower bounds of any of the confidence intervals included 1.0. Thus, there is mixed evidence on the discrimination between the EXT-IJ and ID-INT factors.

The primary support for concluding that the twenty-item model is the better model is based on the improved fit indices and the increase in vari-

TABLE 5
Composite Reliability and Variance Extracted for Five Factor Models

	22-item Model		20-item Model	
	Composite Reliability	Variance Extracted	Composite Reliability	Variance Extracted
AMT	.71	.38	.71	.38
EXT	.79	.44	.80	.49
IJ	.70	.32	.64	.37
ID	.67	.34	.67	.36
INT	.73	.41	.73	.41

ance extracted for the EXT, IJ, and ID factors. While the degrees of freedom have reduced in the twenty-item scale, reduced values of the standardized residuals, improved fit indices, suggest it is the better fitting model. It is also more parsimonious to eliminate the cross-loadings of the two introjected items.

Discussion

The FTMS-A is a self-report measure of reasons for engaging in free time activities. Examination of each of the motivation subscales indicated that they displayed acceptable measurement properties and reasonable levels of fit. The test of the overall model indicated that without modification the model was minimally acceptable. The deletion of two items from the introjected subscale improved the fit to an acceptable level and provides preliminary evidence of the validity of the FTMS-A scale. However, replication of this finding along with further analysis of the reliability and discriminant validity is needed.

Introjected motivation is a conceptually challenging construct. It represents a form of motivation where self-regulation has been somewhat internalized, yet action is associated with concerns for approval of others, guilt, and esteem-contingent actions. Furthermore, among early adolescents in particular, introjection may be difficult to measure as, in many ways, it is at the crux of autonomy development. As such, it may be a relatively unstable phenomenon (see, for example, Ryan & Connell, 1989). The "I have to" or "I want to" reasons of childhood are transformed as perceptions of the people in adolescents' social worlds (e.g., parents, peers, and coaches/leaders) change. As the young adolescent is afforded greater choice, he or she is also increasingly cognizant of the pressures from others in regard to those choices. As parents encourage adolescents to take responsibility for their actions, youth establish new relationships with them. The pressure for making choices now encompasses "I should" and guilt and obligation enter into the picture. For example, a thirteen year old girl might well choose an activity because she wants to support her mother, not because she really wants to do the activity. This phenomenon, of course, holds true with peers as well. In classroom conversations with many of the youth in this study (surrounding the implementation of the *TimeWise* program), they rarely felt that their peers influenced them. Only after probing and more in-depth discussion did they see that this type of motivation does exist and does fit into their lives. Therefore, the conceptual challenge seems largely influenced by the developmental tasks of the age.

Given the above discussion, it is not surprising that the items in the introjection subscale that did not fit as well were more competence and skill based (I want people to think I'm good and To earn rewards and trophies) rather than those that remained. The remaining items (I will feel badly if I don't, To impress my friends, and I want people to like me) all seem to reflect the "I should" part of introjected motivation. The competence items did not invoke the self-contingent, ego-involved aspects of introjection.

In sum, while the three-item Introjected subscale is the weakest in terms of reliability and factor loadings among the other subscales and overall FTMS-A, it is minimally acceptable. It may be possible to develop new items that better capture the competence-based nature of introjected motivation. In addition, it will be interesting to see how well this subscale measures introjected motivation among an older group of adolescents. Gender differences will also be important to assess.

The ability to measure motivation in a general free time context allows for a number of conceptually and/or developmentally interesting questions to be asked. One, which is one of the reasons we worked on developing this measure, is to examine the role different forms of motivation, or different motivation patterns, have on engaging in problem behavior (or prosocial and healthy behavior). Another is to examine how adolescents' motivation patterns or strength of motivation forms change over time. That is, do adolescents become less intrinsically motivated and more identified as they grow increasingly cognizant of self-regulatory aspects of daily life? A related question is to determine how internalization can be facilitated within the free time domain. What experiences influence an adolescent who reports high levels of amotivation to become more intrinsic or more identified? Again, gender differences, and other person and contextual differences should be taken into account in answering these questions.

From a more practical perspective, the FTMS-A appears to be a useful measure of motivation for adolescents in many settings. Whether it has utility in more "clinical" settings, or in settings where youth are characterized as having single or multiple problem behaviors is unknown. There are two ways that the scale might be used to produce an overall score of motivation. Some researchers have opted to create an overall index score created by weighting the individual scores on each of type of motivation (Goudas, Biddle, & Fox, 1994; Ryan & Connell, 1989). In this type of analysis, focus is on overall relative autonomy. A low score, for example, might indicate an adolescent who lacked internal forms of regulation and was susceptible to peer pressure or needed parents or others to prod or push him or her into an activity.

Another approach to using this type of scale is to use cluster analysis to create motivational profiles (Wang & Biddle, 2001) that illustrate groups that vary in interindividual differences. For example, in preliminary analysis with the FTMS-A we have created a typology of motivations that includes four categories: intrinsics, pleasers, moderates, and apathetics (Caldwell, Baldwin, Smith, & Boone, 2002). Although only preliminary, membership in the more autonomous clusters (intrinsics, pleasers) predicted more positive leisure (e.g., good planning and decision making skills and less boredom) and membership in less autonomous clusters (moderates and apathetics) predicted less positive leisure. Such a multivariate representation of individual functioning is in line with the dominant worldviews in developmental psychology (Cairns, Bergman, & Kagan, 2000). This strategy seems to be promising but needs much work in terms of being useful in any type of diagnostic application. Theoretically, creating motivational profiles using the FTMS-A seems promising.

We have some confidence in the twenty-item FTMS-A. More work needs to be done to establish validity and reliability, but as an initial attempt at measuring this complex phenomenon, the scale and its subscales hold promise. Future validation studies are needed to replicate the findings reported here and to assess the convergent and discriminant validity of the FTMS-A with other established measures.

References

Arbuckle, J. L., & Wothke, W. (1999). *Amos user's guide version 4.0.* Chicago: Smallwaters.

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

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming.* Mahwah, NJ: Lawrence Erlbaum.

Cairns, R. B., Bergman, L. R., & Kagan, J. (Eds.). (2000). *Methods and models for studying the individual: Essays in honor of Marian Radke-Yarrow.* Thousand Oaks, CA: Sage.

Caldwell, L. L., Smith, E. A., Swisher, J., & Vicary, J. (in progress). *TimeWise:* Learning Lifelong Leisure Skills. Developed through funding by the U.S. Department of Health and Human Services under Award No. 1R21 DA13193-01.

Caldwell, L. L., Baldwin, C. K., Smith, E. A., & Boone, T. (2002). *TimeWise*—Learning Lifelong Leisure Skills: Conceptualization and preliminary evidence for a new substance abuse prevention program. Pacific Rim Conference on Leisure Education. January, Honolulu, HI, January 2002.

Chatzisarantis, N. L. D., Biddle, S. J. H., & Meek, G. A. (1997). A self-determination theory approach to the study of intentions and the intention-behaviour relationship in children's physical activity. *British Journal of Health Psychology, 2,* 343-360.

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

Frederick, C. M., & Ryan, R. M. (1995). Self-determination theory in sport: A review using cognitive evaluation theory. *International Journal of Sport Psychology, 26,* 5-23.

Goossens, L. (2001). Global versus domain-specific statuses in identity research: A comparison of two self-report measures. *Journal of Adolescence, 24,* 681-699.

Goudas, M., Biddle, S., & Fox, K. (1994). Perceived locus of causality, goal orientations, and perceived competence in school physical education classes. *British Journal of Educational Psychology, 64,* 453-463.

Hatcher, L. (1994). *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling.* Cary, NC: SAS Institute Inc.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105,* 156-166.

Nevitt, J., & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling, 8,* 353-377.

Pelletier, L. G., Fortier, M. S., Vallerand, R. J., Tuson, K. M., & Briere, N. M (1995). Toward a new measure of intrinsic motivation, extrinsic motivation, and amotivation in sports: The sport motivation scale (SMS). *Journal of Sport & Exercise Psychology, 17,* 35-53.

Pelletier, L. G., Vallerand, R. J., Green-Demers, I., Blais, M. R., & Briere, N. M., (1996). Construction et validation d'une mesure de motivation intrinseque, de motivation extrinseque et d'amotivation vis a vis des activities de loisirs: I'Echelle de motivation vis a vis les loisires (EML) [Construction and validation of the Leisure Motivation Scale]. *Loisir et Société, 19,* 559-585.

Rigby, C. S., Deci, E. L., Patrick, B. C., & Ryan, R. M. (1992). Beyond the intrinsic-extrinsic dichotomy: Self-determination in motivation and learning. *Motivation and Emotion, 16,* 165-185.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57,* 749-761.

Ryan, R. M, & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55,* 68-78.

Tabachnick, B. G., & Fiddell, L. S. (2001). *Using multivariate statistics, (4th ed.).* Needham Heights, MA: 02494.

Vallerand, R. J. (1997). Toward a hierarchical model of intrinsic and extrinsic motivation. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology, 29,* 271-360.

Vallerand, R. J., & Fortier, M. S. (1998). Measure of intrinsic and extrinsic motivation in sport and physical activity: A review and critique. In J. L. Duda (Ed.), *Advances in sport and exercise psychology measurement* (pp. 81-101). Morgantown, WV: Fitness Information Technology.

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., & Vallieres, E. F. (1992). The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement, 52,* 1003-1017.

Vlachopoulos, S. P., Karageorghis, C. I., & Terry, P. C., (2000). Motivation profiles in sport: A self-determination theory perspective, *Research Quarterly for Exercise and Sport, 71,* 387-397.

Wang, C. K. J., & Biddle, S. J. H. (2001). Young people's motivational profiles in physical activity: A cluster analysis. *Journal of Sport & Exercise Psychology, 23,* 1-22.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage.

Willis, G. B. (March, 1994). *Cognitive Interviewing and Questionnaire Design: A Training Manual.* Working Paper Series, No. 7. Office of Research and Methodology, National Center for Health Statistics, CDC, US Dept of HHS.

Zhu, W. (1997). Making bootstrap statistical inferences: A tutorial. *Research Quarterly for Exercise and Sport, 68,* 44-55.