Construct validation of the Behavior and Instructional Management Scale

Nancy K. Martin*, Daniel A. Sass

University of Texas, San Antonio, USA

ABSTRACT

Beliefs related to classroom management vary among teachers and play an important role in classrooms. Despite the importance of this construct, valid measures have proven difficult to develop. This study evaluated the psychometric properties of the Behavior and Instructional Management Scale (BIMS), a short but valid measure of teachers' approaches to behavioral and instructional management. Results revealed a two-factor solution that possessed a good model fit, with large estimated factor loadings using confirmatory factor analysis. Evidence of validity was obtained with the Ohio State Teacher Efficacy Scale. Internal consistency for both subscales was adequate. Implications for future research are discussed.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Classroom management has consistently been identified as a salient concern for teachers (Ladd, 2000; Willower, Eidell, & Hoy, 1967). In addition, a meta-analysis of 50 years of research concluded that classroom management is a powerful component of the overall classroom climate as it impacts the level of student engagement, the frequency of inappropriate behavior, and, by extension, the quality of student learning (Wang, Haertel, & Walberg, 1994).

Although often used interchangeably, classroom management and discipline are not synonymous. The term discipline typically refers to the structures and rules describing the behavior expected of students and teacher efforts to ensure that students comply with those rules. However, the literature generally defines classroom management as an umbrella term that encompasses teacher efforts to oversee the activities of the classroom including student behavior, student interactions and learning (Burden, 2000; Evertson & Harris, 1999; Evertson & Weistein, 2006; Good & Brophy, 2000; Iverson, 2003). It is operationalized as behavioral tendencies that teachers utilize to conduct daily instructional activities. These tendencies reflect the teacher's discipline, communication, and instructional styles. All of these aspects manifest in the teacher's preferences and efforts to attain desirable educational goals. Still,

there is no general consensus regarding the specific facets of the construct.

Teachers' beliefs and attitudes regarding the nature of student behaviors and how to manage classrooms vary and can play an important role in the determination of their behavior (Erden & Wolfgang, 2004; Smart, 2009; Urich & Tobin, 1989; Willower et al., 1967; Wolfgang & Glickman, 1980, 1986). As teachers form opinions about how classrooms should be managed, they attempt to operationalize their beliefs by institutionalizing a code for classroom interactions and behaviors (Smart, 2009; Urich & Tobin, 1989). Observation of classroom teachers is one excellent way to gain information regarding classroom management beliefs and practices. However, research suggests other means may also be appropriate and provides evidence of a relationship between teachers' behaviors and their beliefs. Urich and Tobin (1989) describe a first-year teacher's ideological evolution from "teacher as comedian" "to" "teacher as social director." This shift in thinking resulted in changes in teaching behavior and striking improvements in the classroom environment. Similarly, Smart's (2009) observational study triangulated teacher interviews regarding beliefs with classroom behaviors. Thus, it seems feasible that a link exists between teacher beliefs and perceptions regarding classroom management style and proclivity to behavior.

Clearly, the study of this construct has important implications for the dynamics of teaching and learning. Although a large corpus of research exists regarding discipline, the body of knowledge related to the broader construct of classroom management is much smaller. Research in this area has been stymied by the complexity of the

* Corresponding author at: The University of Texas-San Antonio, Department of Educational Psychology, One UTSA Circle, San Antonio, TX 78249, USA.
E-mail address: nancy.martin@utsa.edu (N.K. Martin).

construct and the nature and quality of instruments presently available to measure it. In order to create significant momentum in this field of research, a quality instrument is needed. To that end, the purpose of this paper is to present a new measure of the construct.

1.1. Nature and quality of existing measures

Historically, research has relied on the widespread use of two instruments, the Pupil Control Ideology (PCI) (Willower et al., 1967) and the Beliefs on Discipline Inventory (BDI) (Glickman & Tamashiro, 1980; Wolfgang & Glickman, 1980, 1986). Both scales are focused on the narrower concept of discipline rather than the broader construct of classroom management.

Similarly, in an effort to measure teachers' efficacy related to classroom management, a number of scales include a dimension that attempts to address this facet of the construct, e.g.: Classroom Management/Discipline Efficacy, and the Ohio State Teacher Efficacy Scale (De la Torre Cruz & Casanova Arias, 2007; Emmer & Hickman, 1991; Tschannen-Moran & Woolfolk Hoy, 2001). However, these instruments address teachers' perceptions regarding their ability to create an orderly classroom rather than their approach to it. Although one's efficacy and classroom management approach are likely related, the two are not the same. This premise further demonstrates the need for a behavior management and instructional management scale, as the approach taken purportedly depends on both the teachers' attitudes and the classroom environment. Stated differently, what a teacher believes is the best behavior and instructional management style may not be realized depending on the class environment.

Based on the Beliefs on Discipline Inventory (Wolfgang & Glickman, 1980), Martin and colleagues developed the Attitudes and Beliefs on Classroom Control Inventory (ABCC) (Martin, Yin, & Baldwin, 1998; Martin, Yin, & Mayall, 2007), formerly titled the Inventory of Classroom Management Style (ICMS) (Baldwin & Martin, 1994). Although this instrument addresses the broader construct of classroom management, neither the original nor revised versions are without psychometric concerns. One major problem with this measure was the very high interfactor correlation, thus making it inappropriate for factor analysis models and lacking discriminate validity. In addition, researchers often cannot use these subscales in combination given that multicollinearity often existed and thus neither predictor would provide any unique variability. Also troubling, is the fact that items often did not load on the theorized factor structure using an exploratory factor analysis (EFA) (Henson, 2003; Savran & Cakiroglu, 2003). Clearly, a more refined instrument is needed.

1.2. Theoretical framework

How teachers interact with students is based on their personal sets of beliefs regarding how children develop (Erden & Wolfgang, 2004). The teacher's objectives and approach will vary depending on the theoretical lens through which he or she views their students. Glickman and Tamashiro (1980) and Wolfgang (1995) conceptualized a framework to explain teacher beliefs regarding child development. Based on an integration of theoretical perspectives, the underlying continuum of control underlies the dimensions within the BIMS and hypothesizes three approaches to teacher–student interaction: non-interventionist, interventionist, and interactionalist.

The non-interventionist assumes the child has an inner drive that needs to find its expression in the real world. Interventionists anchor the opposite end of the continuum and emphasize what the outer environment does to shape the human organism in a particular way. The non-interventionist is the least directive and controlling, while the interventionist is most controlling. Traditional behavior modification provides the theoretical foundation for the interventionist's

school of thought. Models of classroom management such as those developed by Canter (1992) or Jones (Jones, 1987; Jones & Jones, 1990) are examples of the interventionist approach. Proponents of Berne (1964), Kohn (1996), Harris (1967) (transactional analysis), Ginott (1972) (congruent communication), or Gordon (1974) (teacher effectiveness training) are considered non-interventionists.

Midway between these two extremes, interactionalists focus on what the individual does to alter the external milieu, as well as what the environment does to shape the individual. Interactionalists work to find solutions acceptable to both the teacher and students and use some of the same techniques as both non-interventionists and interventionists. Theories developed by Adler, Dreikurs, and Glasser provide the framework for interactionalist ideology (Wolfgang, 1995). Cooperative Discipline (Albert, 1989) and Judicious Discipline (Gathercoal, 1990) are examples of classroom management models based on interactionalist ideology. While it is assumed that teachers believe and act according to all three approaches, one usually predominates (Wolfgang, 1995; Wolfgang & Glickman, 1980).

1.3. The connection between behavior management and instruction

It makes sense that the teacher's approach to instruction would be related to their methods of behavior management (Woolfolk Hoy & Weinstein, 2006). For example, one would expect direct instruction to be accompanied by a focus on rules, repetition of academic skills to be coupled with expectations of obedience. Conversely, student-focused instruction such as discussion and active inquiry present higher activity and noise levels in the classroom and result in different behavior management challenges.

In practice, however, the connection between instruction and behavior management may be inconsistent. While it is widely accepted that teachers vary in their classroom approaches, individual teachers may also vary within themselves. For example, teachers who focus on constructivist approaches to instruction may simultaneously emphasize strict adherence to rules. As McCaslin and Good (1998) explain, "Educators have created an oxymoron: a curriculum that urges problem solving and critical thinking and a management system that requires compliance and narrow obedience" (p. 73). While there is little research to document or dispute this mismatch, Garrett's (2006) qualitative study lends some evidence that a lack of connection between how teachers' think about instruction versus behavior management may exist. While teachers thought about instruction as teacher/student-centered, they did not view behavior management through the same lens.

The overarching goal of this paper is to evaluate a new instrument, the Behavior and Instructional Management Scale (BIMS). The BIMS is a relatively brief, posited psychometrically sound instrument that measures teachers' perceptions of their approaches to both behavior management and instructional management. As indicated above, other measures assess either teacher beliefs or teacher efficacy regarding behavior and instructional management, but to our knowledge a psychometrically sound instrument that evaluates teacher perceptions of their actual classroom behaviors is unavailable. Therefore, this measure is critical to the study of differences that may exist between one's beliefs and the ability to execute them within the classroom.

2. Methods

2.1. Participants

Data were collected online from 550 certified teachers employed by three (two urban and one rural) public school districts in the southwestern United States. Most participants were females (81.6%)

and employed by urban school districts (90.0%). The participant pool was composed of 3.3% African-American, 1.1% Asian, 61.4% Caucasian, 31.6% Hispanic, .9% Native American, and 1.8% biracial. Participants ranged in age from 23 to 66, with an average age of 41.82 years ($sd = 10.87$).

The majority of teachers taught either elementary (55.3%) or middle (33.0%) school, with only 10.6% at the high school level. A few teachers (1.1%) taught a mixture of elementary, middle, and/or high school. Teachers taught classes across the curriculum, including both required and elective courses, with an average class size of 23.2 students ($sd = 15.67$, Q1 = 20, Q3 = 25). Most teachers earned either Bachelors (41.3%) or Masters (35.6%) degrees, with a few earning a Doctorate (.4%). A small percentage of teachers reported being enrolled in or taking some Masters (19.4%) or Doctorate (3.3%) level classes. Years of experience ranged from one to 44, with a mean of 13.51 years ($sd = 9.59$). Nearly all teachers were certified (98.7%) and teaching in their certification area (99.3%), with most (83.7%) receiving their teacher certification from a traditional university preparation program. Finally, nearly all teachers (87.3%) reported that they received classroom management training and/or instruction within the past five years.

2.1.1. Measures

2.1.1.1. Theoretical development of the Behavior and Instructional Management Scale (BIMS).
Within this study, classroom management style is defined as a multi-faceted construct that includes two independent constructs: behavior management and instructional management. The continuum of control posited by Glickman and Tamashiro (1980) and Wolfgang (1995) provides the theoretical foundation for each of the two components. Even though teachers fall on this range, their places may vary with regard to managing instruction and classroom behaviors.

Behavior management (BM) is similar to, but different, from discipline in that it includes pre-planned efforts to prevent misbehavior as well as the teacher's response to it. Specifically, this facet includes establishing rules, forming a reward structure, and providing opportunities for student input. Emmer, Evertson, and Anderson (1980) documented one of the primary differences between effective and ineffective classroom managers was the manner in which they formulated and implemented classroom rules. Still, classroom rules are of little assistance if students are not motivated to follow them. As Evertson and Weistein (2006) explain, "… how a teacher achieves order is as important as whether a teacher achieves order" (p. 4). Establishing an effective reward structure and encouraging student input can be useful tools in the prevention of misbehavior and the maintenance of order in the classroom environment.

Instructional management (IM) addresses teachers' instructional aims and methodologies and includes aspects such as monitoring seatwork and structuring daily routines as well as the teacher's use of lecture and student practice versus interactive, participatory approaches to instruction. To what degree does the teacher encourage students to actively interact in the classroom? When designing lessons, to what extent does he or she consider the nature of students − their interests, needs, and background?

The manner in which the teacher approaches instructional tasks contributes to the general classroom atmosphere and classroom management style (Burden, 1995; Kounin, 1970; McNeely & Mertz, 1990; Reeve & Jang, 2006; Weinstein & Mignano, 1993). Nowhere is this better documented than in Kounin's classic (1970) study of orderly and disorderly classrooms. Concepts such as smoothness and momentum of instruction were consistently found to be characteristics of well-planned lessons that prevented off task behaviors. Further, McNeely and Mertz's (1990) study revealed that student teachers began their field experience by focusing on quality lesson planning. By the end of their experience, however, they had begun to see students as the "enemy" and shifted the focus of lesson planning from activities designed to encourage learning to those likely to discourage disruption. More recently, Reeve and Jang (2006) revealed that instructors supportive of student autonomy differ in the types of instructional behaviors from more controlling teachers.

2.1.1.2. Construction of the BIMS.
The BIMS was developed in five stages to create the subscales of Behavior Management and Instructional Management. First, operational definitions for the hypothesized dimensions were developed. Second, a large set of items was generated based on these operational definitions and existing literature, as well as classroom expertise and observations. Third, students enrolled in a graduate course titled Classroom Management and Motivation were surveyed and asked to determine the clarity and content validity of each item on a six-point scale ranging from (1) "not at all" to (6) "very well/very clear." In addition to the operational definitions to rate content validity, students were also asked to supply written feedback for any items that were either unclear or unrelated to the constructs. Fourth, items were revised based on student feedback and pilot tested using a small sample ($n = 94$) of K-12 teachers enrolled in a variety of graduate level courses. Using this small sample, preliminary exploratory factor analyses and reliability analyses were conducted. This information was used to modify those items with poor estimated factor pattern loadings or those that reduced the measure's internal consistency. Items that exhibited limited variability were either revised or removed from the instrument. Finally, to re-evaluate those items with limited variability, the instrument was pilot tested again with a small ($n = 36$) sample of K-12 classroom teachers.

At the conclusion of this five-step process, the BIMS (see Appendix) was composed of two subscales with 24 items underlying the proposed classroom management constructs: Behavior Management (12 items) and Instructional Management (12 items). A 6-point response scale from "strongly agree" to "strongly disagree" was utilized and scoring for some items was reversed. A score for each subscale is obtained by averaging responses across all items or by means of factor analysis.

According to the continuum originally suggested by Wolfgang and Glickman (1980, 1986), endorsement of an item reflects the degree of control the teacher asserts over students. High subscale scores indicate a more controlling, interventionist approach while lower scores are indicative of a less controlling belief in that dimension of classroom management style.

2.1.1.3. Ohio State Teacher Efficacy Scale (OSTES, Tschannen-Moran & Woolfolk Hoy, 2001).
To provide convergent and discriminate validity evidence for the BIMS, teacher data were also collected on the OSTES. Teachers' sense of efficacy, defined as, "… teachers' judgments about their abilities to promote students' learning" (Woolfolk Hoy & Spero, 2005), can have significant impact in the classroom thus relationships with behavior and instructional management are likely. To increase specificity, Tschannen-Moran and Woolfolk Hoy (2001) explained the construct (based on Gibson & Dembo, 1984 research) to include three dimensions of teacher efficacy measured by the OSTES.

The constructs are Efficacy for Instructional Strategies, Efficacy for Classroom Management, and Efficacy for Student Engagement. The Instructional Strategies factor addresses the teacher's perceived ability to tailor instruction to meet student needs and includes aspects, such as gauging learning vis-à-vis their instructional approach and questioning techniques. The Classroom Management factor assesses the teacher's efficacy related to both preventive and reactive attempts to control student behavior. The Student Engagement factor focuses on the teacher's efficacy to

foster support for student learning and motivate all students, including difficult and struggling learners. The short and long versions of the OSTES assess these three constructs with 4- and 8-items per factor, respectfully.

Psychometric research conducted by Tschannen-Moran and Woolfolk Hoy (2001) provided reasonably good evidence of validity and reliability. An exploratory factor analysis with a Varimax rotation revealed a 3-factor solution for both the short and full versions, along with a single higher-order factor of teacher efficacy. This higher-order factor was expected given the high interfactor correlations ($r > .60$) within their study.

Evidence of concurrent validly was revealed as the OSTES subscales were correlated with scales on the Teacher Efficacy Scales (Hoy & Woolfolk, 1990) and the Rand Items. The short and long versions of the OSTES were also highly correlated across the three subscales ($r > .95$). Tschannen-Moran and Woolfolk Hoy (2001) reported relatively high estimated internal consistency (Cronbach's α) coefficients for both the short ($\alpha's > .80$) and long ($\alpha's > .85$) versions of the OSTES. The estimated coefficients using our data for the long and short versions were .87 and .78, .92 and .88, and .90 and .85 on the Efficacy for Instructional Strategies, Efficacy for Classroom Management, and Efficacy for Student Engagement, respectively.

2.2. Procedures

Data were collected online from certified teachers employed at six high schools (five urban and one rural) as well as the middle and elementary schools that feed into them. Based on the assumption that teachers may differ depending on the school context, data were drawn from a thorough cross-section of school clusters chosen based on their socio-economic and ethnic profiles. Principals at the selected campuses were sent an email asking them to forward an invitation to participate to their teachers. The email was sent to 81 principals (57 elementary, 17 middle school, & 7 high school). The email included a link to the consent form and surveys. A follow up reminder was emailed approximately one week later to each principal. Unfortunately, a response rate could not be calculated given that there was no way to determine how many teachers actually received the email.

3. Statistical analyses

3.1. Missing data

A central component to data analysis is the consideration of missing data and how best to treat it. Rubin (1976) proposed three types of missingness: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Statistical analyses are largely robust to missing data when the amount of missingness is minimal (in terms of percent of missing data) or MCAR (Muthén, Kaplan, & Hollis, 1987). As the amount of missingness increases or data are MAR or MNAR, traditional missing data methods (e.g., listwise deletion, pairwise deletion, simple imputation, mean imputation, etc.) often produce biased parameter estimates and reduced statistical power (Enders, 2001).

If data are MCAR or MAR, modern procedures, such as full information maximum likelihood (FIML) estimation and multiple imputations, can ameliorate estimation bias and increased statistical power. Although the percent of missing data was minimal (1.1%), all missing data were treated using the FIML estimation procedure within Mplus given that data were assumed to be MAR. Regardless of how the data are missing (i.e., MAR or MNAR), FIML performs better than most other missing data methods.

3.2. Model fit

The statistics employed to evaluate model fit for exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were the χ^2, Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), Standardized Root Mean Square Residual (SRMR) or Weighted Root Mean Square Residual (WRMR), and Root Mean Square Residual (RMSEA). Description of these model fit statistics can be obtained from Bollen (1989), Hoyle (1995), and Hu and Bentler (1999).

It is well known that χ^2 statistics are nearly always large and statistically significant for complex models. Moreover, the χ^2 statistic is sensitive to large sample sizes and departures from multivariate normality and, therefore, may reject well fitting models. For these reasons, less emphasis was placed on the χ^2 statistics compared to the other model fit statistics (e.g., TLI, CFI, SRMR/WRMR & RMSEA). In general, CFI and TLI statistics greater than .90 were considered as an "adequate" model fit, whereas values greater than .95 were deemed as a "good" model fit (Hu & Bentler, 1999). They denote fit indexes for RMSEA and SRMR values less than .08 and .06, respectively, as "good" and values between .08 and .10 as "mediocre". Acceptable values for WRMR, which are used with the CFA models, are currently unavailable. However, values closer to one are desirable. Note that the CFI, TLI, RMSEA, and SRMR cut points are only for validated for CFA models, as they have not been formally determined for EFA. Instead, CFI and TLI values closer to one and SRMR and RMSEA closer to zero were considered better for EFA models.

4. Instrument evaluation

The BIMS was analyzed in three separate studies. In Study 1, the original 24 items were reduced to 12 items using an EFA with a smaller subsection of the sample. For Study 2, the remainder of the sample was used to evaluate the psychometric properties of the 12-item BIMS using CFA. Study 2 also obtained reliability estimates for each subscale. Finally, Study 3 addressed the discriminate and convergent validity of the 12-item BIMS.

4.1. Study 1

4.1.1. Purpose of study 1

The purposes of Study 1 were twofold: 1) provide preliminary evidence of the psychometric properties of the 24-item BIMS and 2) reduce the number of items to create a 12-item version (6 for BM & 6 for IM) with acceptable psychometric properties. More specifically, the first set of analyses was conducted to determine those items that "best" measure each construct. In essence, this study followed the recommendation of DeVellis (2003, p. 63) by creating more items than intended for use. This step is critical, as past research has produced extremely high interfactor correlates, which are not conducive to factor analysis (see Gorsuch, 1983) and provides little unique variance in later statistical analyses.

4.1.2. Instrument length

DeVellis (2003, p. 66) indicated that the item pools are often three or four times larger than the desired final instrument length (e.g., academic tests), but at times as small as 50% larger for constructs that are difficult to measure (e.g., psychological constructs). For this study, twice as many items were written than were actually intended for use given that the behavior and instructional management constructs are often difficult to assess (or there are only so many different behaviors that can be measured) as reported in the Introduction. While reducing the number of items on an already validated instrument could potentially impact the psychometric properties, this shortcoming is of less concern during instrument development.

Study 1 sought to create a subset of items that "most purely" measure behavior and instructional management. The difficulty with instrument development is determining the optimal number of items necessary to measure each construct, as one must weigh the trade-off between instrument length and reliability. Worthington and Whittaker (2006) advocated for creating instruments that take no more than 15−20 min to administer. Clearly neither our 24- or 12-item instrument would take more than 15 min, but this instrument was developed with the assumption that researchers will likely combine it with other teacher variables (e.g., motivation, efficacy, teaching style, burnout, etc.).

With this literature in mind, 24 items were written (12 per factor) with the intent of retaining the best 12 items (six per factor). Six items per factor was selected for two reasons: 1) Past measures of teacher variables have often found similar lengths appropriate and effective (e.g., Hui & Chan, 1996; McLaney & Hurrell, 1988; Tschannen-Moran & Woolfolk Hoy, 2001) and 2) most importantly, our research revealed that adding additional items did not significantly increase the internal consistency or the factor structure integrity. The latter is critical during instrument development (Worthington & Whittaker, 2006).

4.1.3. Sample sizes for factor analyzes

As indicated later, only 200 subjects were utilized for the EFA, thus leaving 350 for the CFA. Suggested the following sample size guidelines for factor analyses: $100 = $ poor, $200 = $ fair, $300 = $ good, $500 = $ very good, 1000 or more $= $ excellent. However, more recent empirical research (MacCallum, Widaman, Zhang, & Hong, 1999) revealed that the adequacy of factor analysis results depend more on the data characteristics (i.e., communalities) than on the sample size employed. As can be calculated from Table 1 and Fig. 1, the communalities (defined as the sum of the squared loadings for each item) were generally moderate to large in size. Therefore the sample sizes should be adequate for both the EFA and CFA based on the guidelines of MacCallum et al. (1999). In any case, due to its exploratory nature, a smaller sample size ($n = 200$) was selected for Study 1 (i.e., the EFAs) given that the estimated factor pattern

Table 1
Estimated factor pattern loading matrix for the initial and final EFA model.

	Initial EFA model			Final EFA model	
	Factor 1	Factor 2		Factor 1	Factor 2
BM1	*.53	.06	BM1	*.50	.09
BM2	*.76	*−.19	BM2	*.70	−.14
BM3	.06	*.35	BM8	*.58	.15
BM4	*.47	−.10	BM9	*.69	*.20
BM5	−.11	*.68	BM11	*.88	.02
BM6	*.26	−.06	BM12	*.64	−.14
BM7	*.58	*.22	IM2	.17	*.66
BM8	*.59	*.20	IM3	.10	*.65
BM9	*.63	*.27	IM5	.04	*.66
BM10	*.56	*.46	IM6	−.14	*.65
BM11	*.82	.09	IM9	.05	*.56
BM12	*.63	−.10	IM12	−.05	*.83
IM1	*.65	−.17			
IM2	*.18	*.64			
IM3	.03	*.67			
IM4	*.61	.02			
IM5	.05	*.63			
IM6	*−.18	*.66			
IM7	.13	*.60			
IM8	*.53	*.33			
IM9	−.01	*.68			
IM10	*.60	−.16			
IM11	*.50	*−.21			
IM12	−.08	*.78			

Note. Bolded estimated factor pattern loadings reflect items theorized to load together based on content. Factor pattern loadings marked with an "*" were statistically significant at $\alpha^* = .05$.

loading stability was less of a concern compared to the confirmatory portion of this study. Consequently, a larger sample size ($n = 350$) was selected for Study 2 (i.e., the CFA) to ensure more stable and accurate estimates factor loading and model fit statistics.

4.1.4. Statistical analysis with EFA

From the larger sample ($n = 550$), a random sample of participants ($n = 200$) was selected for Study 1. Using the 24-item full scale (12 measuring Behavior Management & 12 measuring Instructional Management), data were analyzed using an EFA within Mplus 5.21 (Muthén & Muthén, 1998−2007). This analysis was performed on a polychoric correlation matrix using a weighted least-squares with mean and variance (WLSMV) estimation procedure with an oblique CF-Equamax rotation. The CF-Equamax rotation, which combines the Quartimax and Varimax criteria (see Browne, 2001), simplifies both variables and factors complexity by spreading the variances more equally across the factors (Gorsuch, 1983). Therefore, items that cross-loaded, or loaded on both factors, could more easily be detected. This rotation criterion is critical given that the factors (i.e., Behavior Management & Instructional Management) are often highly correlated and one aim of this study was to increase discriminate validity and create two reasonably uncorrelated (i.e., $r < .50$) factors.

A two-factor solution was extracted given that items were written for only two factors. Again, the EFA was conducted to 1) shorten the instrument and 2) detect those items that are the "most pure" measure of each factor. For this reason, items were deleted in the following order 1) loaded on the incorrect factor, 2) possessed a higher cross-loading on another factor (defined as having factor pattern loadings $> .30$), and 3) had a small estimated factor loading ($\lambda < .40$) on the theorized factor. Items were removed independently based on the item severity until six items remained per subscale.

4.1.5. Results

The results of the initial EFA model with 24 items indicated a dominant two-factor solution based on the scree plot and eigenvalues, with factors 1 and 2 explaining 24.66 and 17.5 percent of variance, respectively. Additional support for the two-factor model was obtained using the model fit statistics, χ^2 (229) $= 673.350$, $p < .001$, CFI $= .934$, TLI $= .921$, RMSEA $= .098$, SRMR $= .072$, as this model fit is significantly better than the one-factor model, χ^2 (252) $= 2749.352$, $p < .001$, CFI $= .629$, TLI $= .594$, RMSEA $= .223$, SRMR $= .165$. The three-factor model produced a noticeable increase in model fit, χ^2 (207) $= 461.730$, $p < .001$, CFI $= .962$, TLI $= .950$, RMSEA $= .078$, SRMR $= .055$. However, this third factor, which only explained 6.91% of the variance and was not of theoretical interest. Instead, the intent was to remove any items that may be measuring a third factor.

Despite the evidence of a two-factor model, this EFA was only used to shorten the measure and eliminate items that appeared problematic. Using the initial factor loading matrix (see Table 1), the items were removed sequentially based on whether items loaded on the wrong factor (e.g., BM5) and had larger cross-loading magnitudes. In general, several items either had large factor pattern loadings on the incorrect factor or large cross-loadings. This was partially expected given past research with these scales. After each item was removed, another EFA was conducted until there were six items per factor. This approach was employed given that the removal of items influences the other estimated factor pattern loadings.

The final set of items (see Table 1) revealed a two-factor solution with factors 1 and 2 explaining 31.80 and 22.84 percent of variance, respectively. Although factors 1 and 2 had large eigenvalues of 3.816 and 2.741, respectfully, factor 3 only had an eigenvalue of .953 that explained 7.94 percent of the variance. Additional support for the two-factor model was achieved from the model fit statistics,
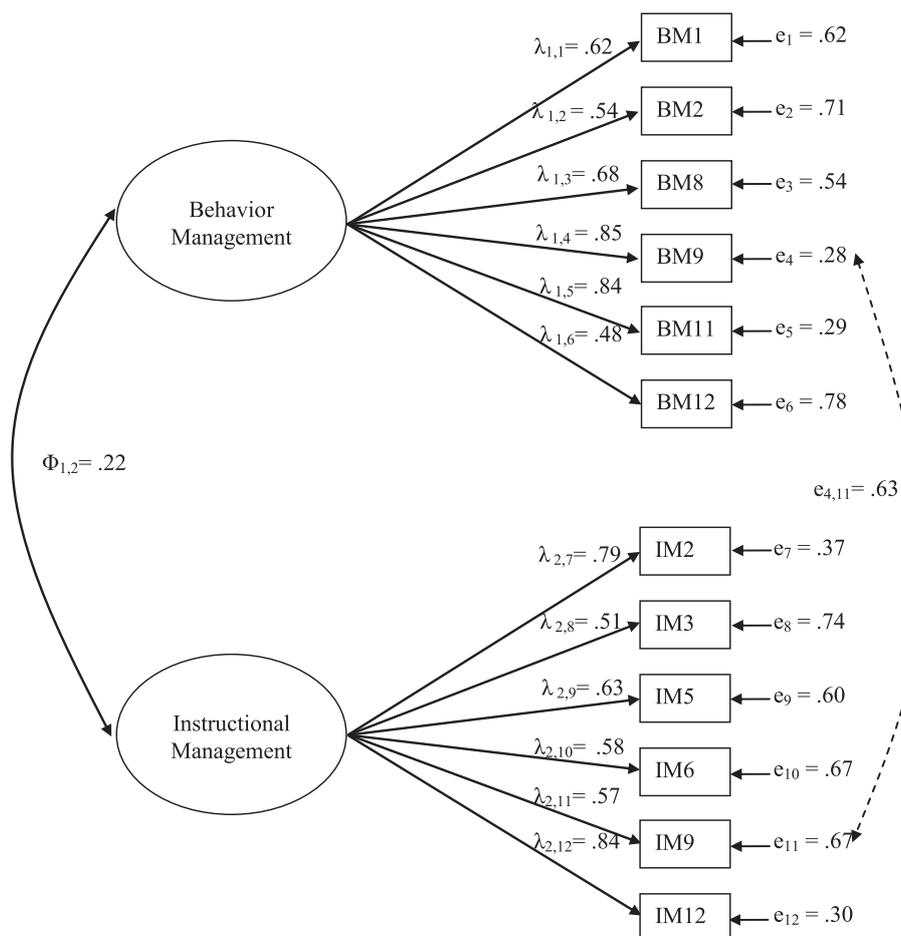
Fig. 1. Estimated factor loadings from the CFA for the shortened version of the BIMS.

χ^2 (43) = 92.153, $p < .001$, CFI = .978, TLI = .966, RMSEA = .076, SRMR = .050, with the three-factor model only providing a modest improvement in model fit, χ^2 (33) = 49.056, $p < .001$, CFI = .993, TLI = .986, RMSEA = .049, SRMR = .034.

To test the estimated factor pattern loadings for statistical significance, the formula $\alpha^* = \alpha/dc$ was used to calculate the new critical value corrected for Type 1 error (Cudeck & O'Dell, 1994). In the equation to compute α^*, $dc = vf - f(f - 1)$, where v is the number of variables (i.e., items), f is the number of factors, and α is the original critical value (i.e., $\alpha = .05$). Therefore, the new two-tailed critical value was $\alpha^* = .001$ ($z_{\alpha^*} = \pm 3.29$) for the initial EFA model and $\alpha^* = .002$ ($z_{\alpha^*} = \pm 3.09$) for the final EFA model.

The estimated factor pattern loadings for the final model (Table 1) revealed that each item had large and statistically significant factor loadings on its hypothesized factor. In addition, the cross-loadings were relatively small (i.e., $\lambda < .20$), and in most cases statistically non-significant. The interfactor correlation was fairly small ($r = .12$), suggesting these factors are measuring unique constructs.

4.1.6. Discussion

The primary objective of Study 1 was to create a reasonably short and useable instrument to measure teacher's perceptions of their behavior and instructional management styles. This is important as participants, in general, are more likely to complete shorter surveys (Dillman, 2000). This shorter version of the BIMS also provided good preliminary factorial validity evidence based on the good model fit and estimated factor pattern loadings. In general, all the estimated factor pattern loadings on the expected factors were large, with relatively small cross-loadings.

Given the nature of Study 1, one major limitation needs to be addressed. Several alternative forms or sets of items could have been evaluated or used to create the shortened version of the BIMS. In other words, selecting items from the larger population of items in many ways may be arbitrary and researchers could use alternative procedures to shorten the instrument. Given that all the items provided good content validity as indicated by content experts and similar distribution characteristics, the choice of using EFA seemed reasonable. Moreover, this approach is commonly employed and recommended within research to shorten measures based on statistical evidence (Coste, Guillemin, Pouchot, & Fermanian, 1997; Worthington & Whittaker, 2006).

The results of Study 1 are promising as the creation of the BIMS is a step toward clarifying the construct of classroom management. Further, this new instrument is broader in scope than both the PCI (Willower et al., 1967) and the BDI (Glickman & Tamashiro, 1980; Wolfgang & Glickman, 1980, 1986). Initial validation evidence also indicates that the BIMS is preferable to the ABCC (Martin et al., 1998; Martin et al., 2007). However, because more evidence regarding the BIMS' psychometric properties is necessary, Study 2 was carried out.

4.2. Study 2

4.2.1. Purpose of Study 2

The purpose of Study 2 was to assess the psychometric properties of the shortened measurement via a separate sample using CFA. The first set of analyses tested for factorial validity using CFA, whereas the second set of analyses investigated the measure's

reliability. While the CFA was conducted using the second subsample (i.e., $n = 350$), the reliability analyses were conducted on the complete sample (i.e., $n = 550$) to provide better parameter estimates of internal consistency for future research.

4.2.1.1. Statistical analysis with CFA. The CFA model was estimated using a polychoric correlation matrix with a WLSMV estimator given that data were categorical. For the CFA model, the first factor loading (i.e., reference indicators) on each factor was set to 1.00 for latent variable scaling and statistical identification. The estimated factor loadings are provided in Fig. 1. All other factor loadings were fixed at zero, with the residuals uncorrelated.

4.2.2. Results
4.2.2.1. Confirmatory factor analysis results. CFA analyses revealed an adequate to good model fit for the proposed model (see Fig. 1), χ^2 $(28) = 126.271$, $p < .001$, CFI = .932, TLI = .949, RMSEA = .100, WRMR = 1.142. For this model all the residuals were assumed to be uncorrelated. Given the model fit was slightly lower than desired for the RMSEA, the modification indices were evaluated and indicated that correlating the residuals for BM9 and IM9 would improve the model fit. Using the DIFFTEST procedure within Mplus (see Asparouhov & Muthén, 2006) to compare nested models using the WLSMV estimation procedure, the results revealed a statistically significant improvement in model fit as a consequence of the correlated residuals, $\Delta\chi^2$ $(1) = 35.092$, $p < .001$, ΔCFI = .013, ΔTLI = .010, ΔRMSEA = −.010, ΔWRMR = −.103. Given these error terms are likely dependent upon one another due to item ordering (i.e., these items followed one another on the measure), these residuals were correlated to yield an overall good to excellent model fit, χ^2 $(28) = 106.637$, $p < .001$, CFI = .945, TLI = .959, RMSEA = .090, WRMR = 1.040.

Its worth noting that the model fit for the entire sample $(n = 550)$ was comparable to the subsample without correlated residuals, χ^2 $(30) = 186.402$, $p < .001$, CFI = .927, TLI = .947, RMSEA = .097, WRMR = 1.342, and when the BM9 and IM9 residuals were correlated, χ^2 $(30) = 162.576$, $p < .001$, CFI = .938, TLI = .953, RMSEA = .092, WRMR = 1.262. As with the subsample of $n = 350$, the model fit using the entire sample did significantly improve with the correlated residuals, $\Delta\chi^2$ $(1) = 32.676$, $p < .001$, ΔCFI = .011, ΔTLI = .006, ΔRMSEA = −.005, ΔWRMR = −.080. However, practically speaking the model fit was not considerably greater based on the CFI, TLI, RMSEA, and WRMR.

In addition to the good overall model fit, estimated factor loadings (see Fig. 1) were all large ($\lambda > .45$) and statistically significant at .004 (.05/12). Collectively, these results suggest that each item is highly correlated with the theorized factor as indicated by the estimated factor loadings, while also uncorrelated with the other factor. In addition, this model produced a two-factor model with a relatively low interfactor correlation, suggesting that these factors capture unique information related to classroom management. The only concern was the correlated residuals, which may have simply occurred due to item order rather than item content. From a content perspective, these items (see Appendix) do not appear to be measuring another factor.

4.2.2.2. Reliability analyses. An analysis of the Behavior Management subscale revealed good internal consistency (i.e., Cronbach's alpha) for the six items ($\alpha = .774$), with an average inter-item correlation of .377 ($sd = .091$). The average corrected item-total correlation for this subscale was .529 ($sd = .071$), which suggests the items have good discrimination. Results for the Instructional Management subscale also showed a good internal consistency for the six items ($\alpha = .770$), with an average inter-item correlation of .365 ($sd = .092$). The average corrected item-total correlation for this subscale was .522 ($sd = .086$), which again suggests good item

discrimination. For both subscales, the internal consistency would not increase with the removal of an item. Overall, these results provide evidence of adequate to good internal consistently measures based on Nunnally's (1978) standards and strong item discrimination.

4.2.3. Discussion
Study 2 evaluated the psychometric properties of the 12-item version of the BIMS and provided promising results. Based on the good model fit for the samples of 350 and 550, these results provide evidence of strong factorial validity given that each indicator (or item) is highly correlated with its corresponding factor and not with the other factor. As seen in Fig. 1 for the $n = 350$ subsample, all the estimated factor loadings were greater than .45 and the modification indices did not indicate that any of the items cross-loaded. Also of significance is the strong discriminate validity, as Behavior Management and Instructional Management are not too highly correlated ($\Phi_{1,2} = .22$). This finding is critical given that past measures (Martin et al., 1998; Martin et al., 2007) of these constructs produced factors so highly correlated that neither provided any unique information, nor could they both be employed in other statistical analyses without creating multicollinearity problems.

The primary reason for the smaller interfactor correlation within this study was the extreme attention given during item writing to ensure that items were only measuring the intended construct. This is critical as both EFA and CFA attempt to minimize cross-loadings, which consequently results in larger interfactor correlations when items measure both factors. Past research with these constructs contained items that sometimes measured both behavioral and instructional management, thus increasing the interfactor correlations and producing poor discriminate validity. By removing items with larger cross-loadings within the EFA, this reduced the interfactor correlations within the CFA (and the EFA).

Although it is very common for researchers to modify model fit using the modification indicated, as Browne (2001) and others suggest, this essentially results in CFA becoming partially exploratory. This may be inappropriate (see Asparouhov & Muthén, 2006; Browne, 2001; Gerbing & Hamilton, 1996; Hurley et al., 1997) as the model modifications may produce a significantly better fitting model simply as a result of a capitalization on chance (Gorsuch, 1983; MacCallum, Roznowski, & Necowitz, 1992). Consequently, it is important for future research to vary item order to determine whether the correlated residuals for BM9 and IM9 occurred simply due to chance, the item ordering, or if in fact these items share common variance unrelated to the factor.

4.3. Study 3

4.3.1. Purpose of Study 3
The purpose of Study 3 was to evaluate the discriminate and convergent validity of the BIMS using the entire sample. Based on conceptual differences, operational definitions, and content, it was hypothesized that Behavior Management would be slightly correlated with Instructional Management, but not with Efficacy for Instructional Strategies or Efficacy for Student Engagement. Further, it was hypothesized that there would be a negative correlation between Behavior Management and Efficacy for Classroom Management (Enochs, Scharmann, & Riggs, 1995; Henson, 2003; Woolfolk & Hoy, 1990; Woolfolk, Rosoff, & Hoy, 1990). Similarly, it was expected that Efficacy for Instructional Strategies, Efficacy for Classroom Management, and Efficacy for Student Engagement would be negatively correlated to Instructional Management, as all three teacher efficacy scales have an instructional component to them (Allinder, 1994; Berman, McLaughin, Bass, Pauly, & Zellman, 1977; Guskey, 1988; Stein & Wang, 1988; Tschannen-Moran & Woolfolk Hoy, 2001).

4.3.2. Results

To investigate discriminate and convergent validity of the BIMS, the entire sample ($n = 476$) of teachers who completed both the BIMS and OSTES were evaluated. Notice that several of the teachers ($n = 74$) did not complete the OSTES and therefore were dropped from these analyses. To take into account measurement error, the correlation matrix (i.e., interfactor correlation or Φ matrix) from the CFA was evaluated. This analysis also allowed for an evaluation of the estimated factor loadings and model fit from a different subsample with the BIMS, while also reevaluating the OSTES's factorial, discriminate, and convergent validity.

The CFA analysis, which paralleled the statistical procedure above with the addition of the OSTES factors, produced a poor to adequate model fit based on the CFI, RMSEA, and WRMR using the full OSTES version, χ^2 (112) = 825.252, $p < .001$, CFI = .877, TLI = .967, RMSEA = .116, WRMR = 1.661. Note that this poor and inconsistent model fit was a direct result of the full OSTES version. With the exception of the TLI, the other model fit statistics were poor when the full OSTES version was analyzed separately from the BIMS, χ^2 (79) = 929.711, $p < .001$, CFI = .801, TLI = .972, RMSEA = .150, WRMR = 1.654. It should be noted that the poor and inconsistent model fit was likely attributable to the extremely large interfactor correlations ($r > .90$) between the OSTES factors.

Given this poor model fit, the short OSTES version was assessed and revealed a good to excellent model fit, χ^2 (34) = 223.497, $p < .001$, CFI = .941, TLI = .988, RMSEA = .108, WRMR = .910. Consequently, the short OSTES version was used to establish the discriminate and convergent validity of the BIMS. Despite this improved model fit, the interfactor correlations were still extremely high ($r > .85$) for the short OSTES version, indicating that discriminate validity was poor.

The results of the CFA model with the short OSTES version and BIMS (see Fig. 2 & Table 2) revealed a good to excellent model fit, χ^2 (75) = 400.701, $p < .001$, CFI = .933, TLI = .971, RMSEA = .096, WRMR = 1.337. Similar to Study 2 using the sample size of 350, correlating the IM9 and BM9 residuals did not significantly improve the model fit from a practical standpoint, but instead only from a statistical perspective, $\Delta\chi^2$ (1) = 23.595, $p < .001$, ΔCFI = .002, ΔTLI = .000, ΔRMSEA = −.002, ΔWRMR = −.017. Therefore, these residuals were not correlated. Instead, a small cross-loading ($\lambda_{3,12} = .23$) between BM2 and the Efficacy for Classroom Management factor produced the greatest change in model fit, $\Delta\chi^2$ (1) = 39.013, $p < .001$, ΔCFI = .013, ΔTLI = .006, ΔRMSEA = −.011, ΔWRMR = −.142, and an overall excellent model fit, χ^2 (77) = 339.562, $p < .001$, CFI = .946, TLI = .977, RMSEA = .085, WRMR = 1.195. Taking this cross-loading into account, the estimated factor loadings and interfactor correlations are provided in Fig. 2 and Table 2, respectively.

In terms of discriminate and convergent validity, these analyses verified our hypothesis that Behavior Management would be slightly correlated with Instructional Management ($r = .24$), but not with Efficacy for Instructional Strategies ($r = −.02$) or Efficacy for Student Engagement ($r = −.12$). Further, our hypothesis that there would be
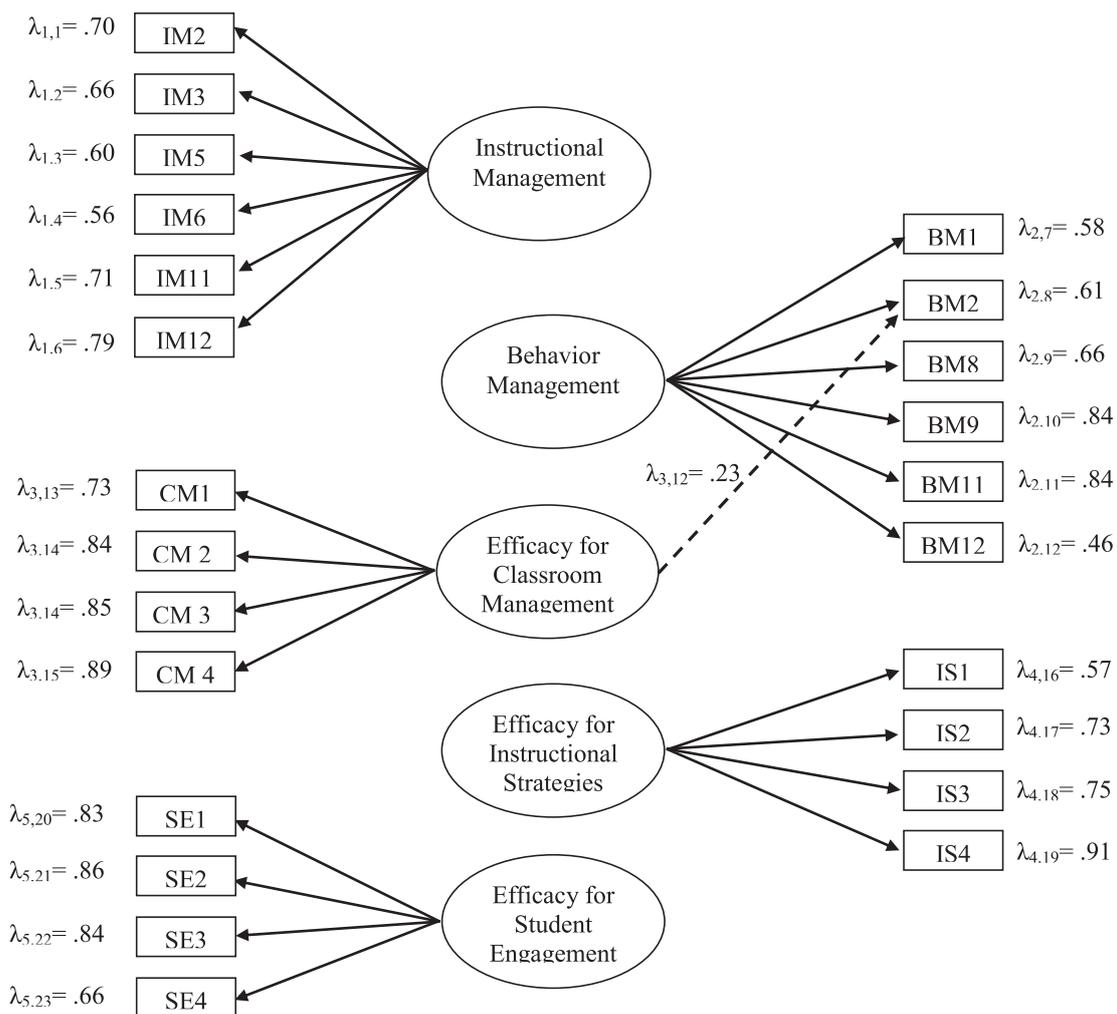

Fig. 2. Estimated factor loadings from the CFA testing the BIMS and OSTES collectively, with the correlations between factors reported in Table 2.

Table 2
Interfactors (Φ's) between the BIMS and teacher efficacy subscales that correspond to the CFA in Fig. 2. Interfactor correlations on the BIMS and OSTES short version.

Measure	BM	IM	IS	CM	SE
BM	–				
IM	**.24**	–			
IS	−.02	**−.59**	–		
CM	**−.19**	**−.51**	**.87**	–	
SE	−.12	**−.65**	**.88**	**.91**	–

Note. BM = Behavior Management, IM = Instructional.
Management, IS = Instructional Strategies, CM = Classroom.
Management, and SE = Student Engagement. Using a Bonferroni adjustment (.05/10), bolded correlations were statistically significant at $p = .005$.

a negative correlation between Behavior Management and Efficacy for Classroom Management ($r = −.19$) was supported, although this correlation was much smaller than anticipated. The strong negative correlations (r's > −.50) between Efficacy for Instructional Strategies, Efficacy for Classroom Management, and Efficacy for Student Engagement and Instructional Management were also confirmed (see Table 2), as all three teacher efficacy scales have an instructional component embedded in them. Overall, these results provide good evidence of discriminate and convergent validity.

4.3.3. A supplemental analysis of the OSTES

Our analyses revealed a large degree of multicollinearity between the OSTES factors, suggesting these scales possess poor discriminate validity (see Table 2) and that perhaps a single measure of efficacy, as measured by the OSTES, is more feasible. To substantiate this finding with the OSTES short and full version, two EFAs using a WLSM estimator and a polychoric correlation matrix with a CF-Equamax rotation were conducted.

Results revealed a one-factor solution based on the scree plot and eigenvalues. The eigenvalues and percent of variance explained for factors 1, 2, and 3 were 7.274 (60.6%), .943 (7.9%), and .695 (5.8%), respectively, for the short version. In addition, the data fit a one-factor model very well, χ^2 (54) = 550.722, $p < .001$, CFI = .984, TLI = .980, RMSEA = .139, SRMR = .051. This finding was replicated for the long OSTES version, with the eigenvalues and percent of variance explained of 13.559 (56.5%), 1.513 (6.3%), and 1.052 (4.4%) for factors 1, 2, and 3, respectively, and a good model fit, χ^2 (252) = 3283.607, $p < .001$, CFI = .972, TLI = .969, RMSEA = .159, SRMR = .064. Although statistically these OSTES subscales provide little unique information, from a content standpoint these three subscales are conceptually distinct. Given that each subscale measures efficacy within the classroom related to instruction, it is not surprising that these subscales are so highly correlated.

4.3.4. Discussion

Study 3 explored the discriminate and convergent validity of the BIMS, with results confirming the hypothesized relationships. As expected, the relationships between the BIMS' two subscales revealed relatively independent relationships, thus providing evidence of discriminate validity. Substantiation of convergent and discriminate validity was also determined by considering relationships of IM and BM to components of the OSTES: instructional strategies, student engagement and classroom management. Results revealed statistically significant inverse relationships between Instructional Management and the components of teacher efficacy (as measured by the short version of the OSTES). Teachers with higher levels of efficacy regarding instructional strategies, student engagement, and classroom management are less likely to take a directive approach in implementing tactics to manage instruction. Conversely, the relationship between behavior management and the three teacher efficacy variables was relatively small. This finding was anticipated given that how a teacher manages student behaviors is

likely unrelated to their perceived ability to effectively manage classroom instruction and keep students engaged.

The above results support previous research in that teacher efficacy is not necessarily strongly related to how a teacher handles classroom management (Henson, 2003; Henson & Chambers, 2005; Hoy & Woolfolk, 1990; Savran & Cakiroglu, 2003; Woolfolk & Hoy, 1990). Specifically, teachers with a high sense of efficacy tended to favor more humanistic and less controlling management orientations (Henson, 2003; Hoy & Woolfolk, 1990; Woolfolk & Hoy, 1990; Woolfolk et al., 1990).

To increase our understanding of the OTES's psychometric properties, additional analyses on both the long and short versions of the OTES were conducted. These results revealed a one-factor solution based on the eigenvalues and model fit statistics, which support the position above that the OTES possesses weak discriminate validity. This conclusion is not entirely different from that of Tschannen-Moran and Woolfolk Hoy (2001), who explained that both versions of the OTES could be used to create a second-order factor called efficacy due to the higher interfactor correlations. However, our results differed from Tschannen-Moran and Woolfolk Hoy (2001) in that our interfactor correlations were noticeably higher. Possible explanations for these correlation differences could be how these correlations were estimated (total raw scores vs. CFA, as CFA corrects for measurement error and weights the items based on the factor loadings), sample characteristics, and/or sample size. Regardless, this study chose to use the subscale scores since these aspects of efficacy vary in concept and in content.

5. Conclusions

The ability to identify, define, and measure the facets of classroom control will provide the means to address a variety of research questions regarding classroom dynamics that, to date, have been largely untapped. The three studies presented here provide evidence for a brief, psychometrically sound instrument designed to measure the aspects of teachers' beliefs toward managing behavior and instruction. Study 1 utilized EFA to examine the 24-item version of the BIMS and reduce it to 12 items. The second and third studies examined the validity (via factorial, discriminate, & convergent validity) and reliability estimates of the shortened version. Collectively, these study provided evidence of adequate psychometric properties. This study also provides validation evidence of the OSTES.

The short version of the BIMS appears to be a promising and emerging measure. However, at this early stage in its development, additional analyses regarding the BIMS' psychometric properties are needed. At this juncture, we recommend that researchers collect data from the 24-item version and determine whether the same 12-item short version manifests. Further, it is also critical that future research assess the optimal number of items to measure behavior and instructional management. For this study, increasing the number of items did not improve the measurement precision (i.e., reliability), but future research may draw a different conclusion with different samples. Moreover, larger sample sizes for the EFA may yield slightly different results and therefore additional research is needed.

Along a similar vein, it is possible that participants' responses to an individual item will be influenced by preceding items. Thus altering the order in which the items are administered is worthy of study. As a result, researchers should consider interspersing BM and IM statements by changing the order the items are administered. Additional information regarding the psychometric properties (e.g., concurrent validity, predictive validity, construct validity, etc.) would bolster the instrument's psychometric foundation.

When considering the construct of teacher control, it is important to recognize that behavior and instructional management styles are probably similar to teacher efficacy in that they are not

static qualities. Rather they are allegedly contextual in nature — a combination of both the person and the environment. For example, research indicates that teacher efficacy may vary depending on a variety of variables, such as cultural backgrounds (Lin, Gorrell, & Taylor, 2002), professional maturity (De la Torre Cruz & Casanova Arias, 2007; Woolfolk Hoy & Spero, 2005), or school setting (urban, suburban, or rural) (Knoblauch & Woolfolk Hoy, 2008). Similarly, teacher perceptions regarding behavior and instructional management should be viewed as qualities that may change over time and situations (i.e., responses may depend on the school, classroom, curricular context, age of students, etc.). Thus, future research should study the stability of these attributes and whether this instrument assesses characteristics of teachers, workplace setting, or both. Accordingly, the authors are not claiming that this instrument measures stable teacher characteristics, but instead only reported behaviors within the current teacher context.

As there are limitations with all research, these studies are no exception. Because of the manner in which data were collected, the return rate cannot be determined. Therefore, these results could be sample specific; teachers who responded may be qualitatively different from the population at large. In other words, the results may be biased if volunteer subjects are fundamentally different from the general population. Factors associated with sampling bias, which could include restriction of range and ultimately biased means and variances, could produce biased factor analysis results. Thus, future research is needed to replicate these results within different sample sizes and characteristics.

A related potential limitation is the method of data collection. These data were collected online to increase efficiency, and because past research has found higher response rates and shorter time frames for the data collection process (Mertler, 2002). Due to the increased use of online data collection, research into online data collection has been examined at length. Findings have been fairly consistent in concluding that the paper–pencil and online surveys yield results that are basically interchangeable (Cronk & West, 2002; Krantz & Dalal, 2000). In addition, research indicates that online data collection is

reliable, valid, and reasonably representative (Meyerson & Tryon, 2003). In any case, future research should test whether the current results remain using a paper–pencil administration of the BIMS.

Further, none of the three studies addressed the important question of invariance across groups (e.g., grade levels, content areas, gender, age, etc). For example, the extent to which the BIMS is an effective measure across elementary, middle and high school levels remains unknown and presents a question pertinent for future studies. Stated differently, testing whether the factor structure (in particular the factor loadings and intercepts) is invariant across grade levels ensures that teachers at these different grade levels interpret the items similarly, thus allowing for accurate mean comparisons. In addition, it is important to note that the BIMS is a self-report instrument that attempts to measure teachers' approaches to classroom control. Therefore, concurrent validity related to observational data is needed to verify this self-report instrument.

Still, the BIMS holds promise, as it appears to be a useful tool for advancing the measurement of this important construct and addressing a variety of important research questions. For example, what are the implications for ushering new teachers into the profession — for preparing pre-service teachers and mentoring novice educators? To what extent are teachers consistent within themselves regarding their approaches to instruction and behavior management? What is the relationship between teacher control and other variables such as teacher stress, burnout, and turn over? Finally, and perhaps most important, what is the nature of the relationship between teachers' approaches to classroom control and student efficacy, stress, achievement, or persistence in school?

There can be little doubt that teachers encounter a variety of experiences in the classroom. Their beliefs regarding these experiences, and the manner in which they approach them, work together to create a unique and individual style of classroom management. A more clear understanding of the construct has important implications for the classroom. To that end, the Behavior and Instructional Management Scale appears to be a timely and worthwhile tool for research regarding this important topic.

Appendix

Directions: For each statement below, please mark the response that best describes what you do in the classroom. There are no right or wrong answers, **so please respond as honestly as possible**.	Strongly agree	Agree	Slightly agree	Slightly disagree	Disagree	Strongly disagree
1. I nearly always intervene when students talk at inappropriate times during class. (BM1)	6	5	4	3	2	1
2. I use whole class instruction to ensure a structured classroom. (IM1)	6	5	4	3	2	1
3. I strongly limit student chatter in the classroom. (BM2)	6	5	4	3	2	1
*4. I nearly always use collaborative learning to explore questions in the classroom. (IM2)	6	5	4	3	2	1
5. I reward students for good behavior in the classroom. (BM3)	6	5	4	3	2	1
* 6. I engage students in active discussion about issues related to real world applications. (IM3)	6	5	4	3	2	1
7. If a student talks to a neighbor, I will move the student away from other students. (BM4)	6	5	4	3	2	1
8. I establish a teaching daily routine in my classroom and stick to it. (IM4)	6	5	4	3	2	1
*9. I use input from students to create classroom rules. (BM5)	6	5	4	3	2	1
*10. I nearly always use group work in my classroom. (IM5)	6	5	4	3	2	1
*11. I allow students to get out of their seat without permission. (BM6)	6	5	4	3	2	1
*12. I use student input when creating student projects. (IM6)	6	5	4	3	2	1
13. I am strict when it comes to student compliance in my classroom. (BM7)	6	5	4	3	2	1
*14. I nearly always use inquiry-based learning in the classroom. (IM7)	6	5	4	3	2	1
15. I firmly redirect students back to the topic when they get off task. (BM8)	6	5	4	3	2	1
16. I direct the students' transition from one learning activity to another. (IM8)	6	5	4	3	2	1
17. I insist that students in my classroom follow the rules at all times. (BM9)	6	5	4	3	2	1
*18. I nearly always adjust instruction in response to individual student needs. (IM9)	6	5	4	3	2	1
19. I closely monitor off task behavior during class. (BM10)	6	5	4	3	2	1
20. I nearly always use direct instruction when I teach. (IM10)	6	5	4	3	2	1
21. I strictly enforce classroom rules to control student behavior. (BM11)	6	5	4	3	2	1
22. I do not deviate from my pre-planned learning activities. (IM11)	6	5	4	3	2	1
23. If a student's behavior is defiant, I will demand that they comply with my classroom rules. (BM12)	6	5	4	3	2	1
*24. I nearly always use a teaching approach that encourages interaction among students. (IM12)	6	5	4	3	2	1

Note. Bolded items represent the 12 items used for the abbreviated scale. Although items are labeled here as BM and IM, these markers did not appear on the version completed by the subjects.
* = item is reverse scored.

References

Albert, L. (1989). *A teacher's guide to cooperative discipline: How to manage your classroom and promote self-esteem*. Circle Pines, MN: AGS.

Allinder, R. M. (1994). The relationship between efficacy and the instructional practices of special education teachers and consultants. *Teacher Education and Special Education, 17*, 86–95.

Asparouhov, T., & Muthén, B. (May 26, 2006). Robust chi square difference testing with mean and variance adjusted test statistics. Mplus Web Notes: No. 10.

Baldwin, B. & Martin, N.K. (1994, April). Using factor analysis to establish construct validity of an Inventory of Classroom Management Style. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA.

Berman, P., McLaughin, M., Bass, G., Pauly, E., & Zellman, G. (1977). *Federal programs supporting educational change. Col. VII: Factors affecting implementation and continuation (Report No. R-1589/7-HEW)*. Santa Monica, CA: The Rand Corporation. (ERIC document reproduction service No. 140 432).

Berne, E. (1964). *Games people play: The psychology of human relations*. New York: Ballantine Books.

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

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*, 111–150.

Burden, P. R. (1995). *Classroom management and discipline*. White Plains, NY: Longman.

Burden, P. R. (2000). *Powerful classroom management strategies: Motivating students to learn*. Thousand Oaks, CA: Corwin Press.

Canter, L. (1992). *Assertive discipline: Positive behavior management for today's classroom*. Santa Monica, CA: Lee Canter & Associates.

Coste, J., Guillemin, F., Pouchot, J., & Fermanian, J. (1997). Methodological approaches to shortening composite measurement scales. *Journal of Clinical Epidemiology, 50*, 247–252.

Cronk, B. C., & West, J. L. (2002). Personality research on the Internet: a comparison of web-based and traditional instruments in take-home and in-class settings. *Behavior Research Methods, Instruments, and Computers, 34*, 177–180.

Cudeck, R., & O'Dell, L. L. (1994). Applications of standard error estimates in unrestricted factor analysis: significance tests for factor loadings and correlations. *Psychological Bulletin, 115*, 475–487.

De la Torre Cruz, M. J., & Casanova Arias, P. F. (2007). Comparative analysis of expectancies of efficacy in in-service and prospective teachers. *Teaching and Teacher Education, 23*, 641–652.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method*. New York: J. Wiley.

Emmer, E. T., Evertson, C. M., & Anderson, L. M. (1980). Effective classroom management at the beginning of the school year. *The Elementary School Journal, 80*(5), 219–231.

Emmer, E. T., & Hickman, J. (1991). Teacher efficacy in classroom management and discipline. *Educational and Psychological Measurement, 51*, 755–765.

Enders, C. K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement, 61*, 713–740.

Enochs, L. G., Scharmann, L. C., & Riggs, I. M. (1995). The relationship of pupil control to preservice elementary science teacher self-efficacy and outcome expectancy. *Science Education, 79*, 63–75.

Erden, F., & Wolfgang, C. H. (2004). An exploration of the differences in prekindergarten, kindergarten, and first grade teachers' beliefs related to discipline when dealing with male and female students. *Early Child Development and Care, 174*, 3–11.

Evertson, C., & Harris, S. (1999). Support for managing learning-centered classrooms: the classroom management and organization program. In H. Jerome Freiberg's (Ed.), *Beyond behaviorism: Changing the classroom management paradigm* (pp. 57–74). Boston, MS: Allyn & Bacon.

Evertson, C. M., & Weistein, C. S. (2006). Classroom management as a field of inquiry. In C. M. Evertson, & C. S. Weinstein's (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 3–15). Mahwah, NJ: Lawrence Erlbaum Associates.

Gathercoal, F. (1990). *Judicious discipline* (2nd ed.). Davis, CA: Caddo Gap Press.

Garrett, T. (2006). Student– and Teacher–Centered Classroom Management: A Case Study of Three Teachers' Beliefs and Practices. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Gerbing, D. W., & Hamilton, J. G. (1996). Viability of exploratory factor analysis as a precursor to confirmatory factor analysis. *Structural Equation Modeling, 3*, 62–72.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: a construct validation. *Journal of Educational Psychology, 76*, 569–582.

Ginott, H. (1972). *Teacher and child*. NY: Macmillan.

Glickman, C. D., & Tamashiro, R. T. (1980). Clarifying teachers' beliefs about discipline. *Educational Leadership, 37*, 459–464.

Good, T. L., & Brophy, J. E. (2000). *Looking in classrooms* (8th ed.). New York, NY: Addison-Wesley Educational Publishers.

Gordon, T. (1974). *Teacher effectiveness training*. New York: Wyden.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Guskey, T. R. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional motivation. *Teaching and Teacher Education, 4*, 63–69.

Harris, T. A. (1967). *I'm OK – You're OK*. New York: Avon Books.

Henson, R. K. (2003). Relationships between preservice teachers' self-efficacy, task analysis, and classroom management beliefs. *Research in the Schools, 10*, 53–62.

Henson, R. K., & Chambers, S. M. (2005). Personality type as a predictor of teaching efficacy and classroom control in emergency certification teachers. *Education, 24*, 261–268.

Hoy, W. K., & Woolfolk, A. E. (1990). Socialization of student teachers. *American Educational Research Journal, 27*, 279–300.

Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage Publications.

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

Hui, E. K. P., & Chan, D. W. (1996). Teacher stress and guidance work in Hong Kong secondary school teachers. *British Journal of Guidance and Counseling, 24*, 199–211.

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

Iverson, A. M. (2003). *Building competence in classroom management and discipline*. Upper Saddle River, NJ: Merrill/Prentice-Hall.

Jones, F. (1987). *Positive classroom discipline*. NY, NY: McGraw-Hill.

Jones, V. F., & Jones, L. S. (1990). *Responsible classroom discipline* (3rd ed.). Boston: Allyn and Bacon.

Kohn, A. (1996). *Beyond discipline: From compliance to community*. Alexandria, VA: Association for Supervision and Curriculum Development.

Kounin, J. S. (1970). *Discipline and group management in classrooms*. NY, NY: Holt, Rinehart & Winston.

Knoblauch, D., & Woolfolk Hoy, A. E. (2008). "Maybe I can teach *those* kids." The influence of contextual factors on student teachers' efficacy beliefs. *Teaching and Teacher Education, 24*, 166–179.

Krantz, J. H., & Dalal, R. (2000). Validity of web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 35–60). San Diego, CA: Academic Press.

Ladd, K. L. (2000). A comparison of teacher education programs and graduate's perceptions of experiences. Dissertation Abstracts International (University Microfilms No. 9998491).

Lin, H., Gorrell, J., & Taylor, J. (2002). Influence of culture and education on U.S. and Taiwan preservice teachers' efficacy beliefs. *Journal of Educational Research, 96* (1), 37–46.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modification in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*, 490–504.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84–99.

McCaslin, M., & Good, T. L. (1998). Moving beyond Management as Sheer Compliance: Helping Students to Develop Goal Coordination Strategies. *Educational Horizons, 76*(4), 169–176.

McLaney, M. A., & Hurrell, J. J. (1988). Control, stress, and job satisfaction in Canadian nurses. *Work and Stress, 2*, 217–224.

McNeely, S. R., & Mertz, N. T. (1990, April). Cognitive constructs of preservice teachers: research on how student teachers think about teaching. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Martin, N. K., Yin, Z., & Baldwin, B. (1998). Construct Validation of The Attitudes and Beliefs on Classroom Control Inventory. *Journal of Classroom Interaction, 33*(2), 6–15.

Mertler, C. (2002). Demonstrating the potential for web-based survey methodology with a case study. *American Secondary Education, 30*, 49–61.

Martin, N. K., Yin, Z., & Mayall, H. (2007). The Attitudes & Beliefs on Classroom Control Inventory-Revised and Revisited: A Continuation of Construct Validation. *Journal of Classroom Interaction, 42*(2), 11–20.

Meyerson, P., & Tryon, W. (2003). Validating internet research: a test of the psychometric equivalence of internet and in-person samples. *Behavior Research Methods, Instruments, and Computers, 35*, 614–620.

Muthén, B. O., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 55*, 107–122.

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

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Reeve, J., & Jang, H. (2006). What teachers say and do to support students' autonomy during a learning activity. *Journal of Educational Psychology, 98*(1), 209–218.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581–592.

Savran, A., & Cakiroglu, J. (2003). Differences between elementary and secondary preservice teachers' perceived efficacy beliefs and their classroom management beliefs. *The Turkish Online Journal of Educational Technology, 2*(4).

Smart, J. (2009, April). A case study of an Italian teacher's behavior management beliefs, practices, and related classroom interactions. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Stein, M. K., & Wang, M. C. (1988). Teacher development and school improvement: the process of teacher change. *Teaching and Teacher Education, 4*, 171–187.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805.

Urich, S. L., & Tobin, K. (March, 1989). The influence of a teacher's beliefs on classroom management. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1994). What helps students learn? *Educational Leadership, 51*, 74–79.

Weinstein, C. S., & Mignano, A. J., Jr. (1993). *Elementary classroom management: Lessons from research and practice*. NY, NY: McGraw-Hill.

Willower, D. J., Eidell, T. L., & Hoy, W. K. (1967). Conceptual framework. *The Pennsylvania State University Studies, 26*, 3–8.

Wolfgang, C. H. (1995). *Solving discipline problems: Strategies for classroom teachers* (3rd ed.). Boston: Allyn and Bacon.

Wolfgang, C. H., & Glickman, C. D. (1980). *Solving discipline problems: Strategies for classroom teachers* (1st ed.). Boston: Allyn and Bacon.

Wolfgang, C. H., & Glickman, C. D. (1986). *Solving discipline problems: Strategies for classroom teachers* (2nd ed.). Boston: Allyn and Bacon.

Woolfolk, A. E., & Hoy, W. K. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology, 82*, 81–91.

Woolfolk Hoy, A. E., & Weinstein, C. S. (2006). Student and teacher perspectives in classroom management. In Carol M. Evertson, & Carol S. Weinstein's (Eds.), *Handbook of classroom Management: Research, practice, and contemporary issues* (pp. 685–709). Mahwah, NJ: Lawrence Erlbaum Associates.

Woolfolk Hoy, A. W., & Spero, R. B. (2005). Changes in teacher efficacy during the early years of teaching: a comparison of four measures. *Teaching and Teacher Education, 21*, 343–356.

Woolfolk, A. E., Rosoff, B., & Hoy, W. K. (1990). Teachers' sense of efficacy and their beliefs about managing students. *Teaching and Teacher Education, 6*, 137–148.

Worthington, R. I., & Whittaker, T. A. (2006). Scale development research: a content analysis and recommendations for best practices. *The Counseling Psychologist, 34*, 806–838.