2010

The Psychometric Properties of the Difficult Behavior Self-Efficacy Scale

Hyun-Kyoung Oh
California State University - San Bernardino

Francis M. Kozub
The College at Brockport, fkozub@brockport.edu

Citation/Publisher Attribution:
Oh, H., & Kozub, F. M. (2010). The Psychometric Properties of the Difficult Behavior Self-Efficacy Scale. Adapted Physical Activity Quarterly, 27(3), 191-207.

Adapted Physical Activity Quarterly, 2010, 27, 191-207

The Psychometric Properties of the Difficult Behavior Self-Efficacy Scale

Hyun-Kyoung Oh
California State University, San Bernardino

Francis M. Kozub
College at Brockport, State University of New York

The study was designed to estimate the psychometric properties of Hastings and Brown's (2002a) Difficult Behavior Self-efficacy Scale. Participants were two samples of physical educators teaching in Korea (n = 229) and the United States (U.S.; n = 139). An initial translation of the questionnaire to Korean and pilot study were conducted along with the larger study using a confirmatory factor analysis procedure. Internal consistency estimates (weighed Omega) for the five-item scale were 0.88 both the Korean and U.S. samples. The average variances extracted for the one factor were 0.59 for the total data set and 0.57 each for the Korean and U.S. samples. Confirmatory factor analysis supported a five-item, unidimensional model for self-efficacy for the total sample: Goodness of Fit Index (GFI) = 0.97, Nonnormed Fit Index (NNFI) = 0.95, Comparative Fit Index (CFI) = 0.98, and Standardized Root Mean Square Residual (SRMR) = 0.03. Only the Root Mean Square Error of Approximation (RMSEA = 0.12) fell below criterion levels of acceptable fit, with similar fit indices occurring in separate analyses of the Korean and U.S. samples. Invariance testing across the two samples supported metric invariance (similarity of factor loadings) but not scalar invariance (U.S. means higher on all five items). The factor structure for the self-efficacy scale provides an initial estimate of validity and internal consistency for use with different teacher groups.

Autism spectrum disorder (ASD) has been recognized for more than 60 years, starting with Kanner (1943), who published the classic autism paper in which he described 11 children with autism. These developmental disorders are characterized by a distinct pattern of social interaction deficit, communication and imagination impairment, and a restricted range of interest (American Psychiatric Association, 1994). Students with ASD often show common types of challenging behavior such as aggressive, self-injurious, and stereotypical behaviors. Challenging behaviors are difficult for both students with ASD and teachers. For students with ASD,

Hyun Kyoung Oh is with the Department of Kinesiology at California State University, San Bernardino. Francis Kozub is with the Department of Kinesiology, Sport Studies, and Physical Education at the College at Brockport, State University of New York.

challenging behaviors create safety as well as learning issues that can impact on peers. From the teacher's point of view, challenging behavior affects classroom climate (Kaiser & Rasminsky, 2003). Teachers often experience difficulty and frustration in managing the challenging behavior of students with ASD, leading to negative perceptions toward integration (Kaiser & Rasminsky, 2003; LaMaster, Gall, Kinchin, & Siedentop, 1998).

Self-efficacy, a belief about personal capacity to successfully perform tasks or actions in a specific situation or context, is a topic of considerable interest among educators who teach learners with disabilities (Bandura, 1997; Hastings & Brown, 2002a; Hastings & Symes, 2002). Researchers believe that self-efficacy is a factor that influences not only how a teacher feels, but also how this individual may respond to challenging situations (Bandura, 1997; Maddux, 1995). An individual's behavior in demanding situations is the crux of why self-efficacy is an important physical education concern. Finding appropriate solutions for dealing with learners who have a history of impairment resulting in challenging behavior is important. Coupling this disability-related issue with the dynamic nature of physical education makes physical educators' self-efficacy an important area that has received little attention in the published literature. Efficacy responses that mediate teacher judgments and expectations for learners as well as their beliefs in personal skills to control situations where challenging behaviors occur is consistent with other applications of self-efficacy theory (Bandura, 1997). Since Bandura introduced self-efficacy theory, many researchers across countless disciplines have studied this construct, thus expanding the scope to include a variety of social, health, athletic, personal psychology, and teaching-related topics (Bandura, 1997; Maddux, 1995).

Self-efficacy beliefs related to teams, performance, and coaching have been a hot research topic in physical activity, exercise and sport sciences (Feltz, 2000; Malete & Feltz, 2000); however, researchers in the physical education field have given little attention to self-efficacy. Existing studies are limited to the impact of mentoring-based professional development on self-efficacy, physical activity-related teaching behavior as a function of teacher self-efficacy, and determinants of teachers' intentions to teach physical activity using self-efficacy as predictor of teacher attitudes (Martin, Kulinna, Eklund, &Reed, 2001; Martin & Kulinna, 2003; Martin & Kulinna, 2005; Martin, McCaughtry, Kulinna, Cothran, & Faust, 2008). In the later study by Martin and Kulinna (2005), it was found that teachers high in self-efficacy in physical education were more likely to teach lessons that were physically active. These teachers were also more likely to overcoming barriers to instruction and have strong intentions as well as positive attitudes toward teaching. Moreover, Martin et al. (2008) found an effect from mentoring-based support on professional development in helping physical education teachers develop efficacy toward using pedometers and computers. Further, physical education teachers' self-efficacy is believed to have an impact on the desire of individuals to remain in teaching jobs long enough to become effective teachers (Martin & Hodges-Kulinna, 2004; Martin et al., 2008). Generalizing self-efficacy responses and outcomes in relation to physical educators and specifically to those responsible for teaching diverse learners is an important research topic.

Teachers of students with autism must understand how their self-efficacy perceptions are related to instructional behaviors that may impact pupil learning; however, researchers have not explored self-efficacy responses toward challenging

learner behavior in physical educational contexts. In general education settings, researchers have indicated that teachers' self-efficacy was an essential factor linked to successful teaching. For example, teachers with higher self-efficacy are more likely to focus on student learning rather than content covered, set higher professional goals, and are more willing to accept responsibility if learner outcomes are not positive (Ross, Cousins, & Gadalla, 1996). This may also be the case for managing learner behavior and utilizing strategies that lead to successful integration of children with disabilities.

Unlike physical education, researchers in other education and special education areas have prioritized teacher efficacy as a variable in need of extensive study including research to develop measures for the construct (Brouwers & Tomic, 2001; Deemer & Minke, 1999; Meijer & Foster, 1988; Wheatley, 2002). Brouwers and Tomic (2001) examined validity of scores on the *Teachers Interpersonal Self-Efficacy Scale*. This instrument contains subscales to measure Managing Student Behavior in the Classroom, Eliciting Support from Colleagues, and School Principals. They indicated that the *Teachers Interpersonal Self-Efficacy Scale* can assess teachers' self-efficacy in general settings. The researchers suggested a need to measure teachers' self-efficacy related to specific tasks, organizational settings, and activities. Meijer and Foster (1988) inferred that self-efficacy is a potential variable that can predict teachers' behaviors related to student referrals, as well as individual attitude and teaching skills in special education settings. Overall, self-efficacy researchers have supported that teachers who have high self-efficacy believe in their abilities to positively impact student outcomes. Finding suitable means to decrease inappropriate acts and stimulate an increase in more socially acceptable behaviors by children with disabilities are desirable outcomes potentially mediated by self-efficacy. In general, teachers are likely to show positive attitudes toward teaching diverse students, a willingness to further develop their teaching skills, and a higher likelihood of adopting teaching innovations when self-efficacy is high (Brouwers & Tomic, 2001; Meijer & Foster, 1988; Wheatley, 2002).

Further, Hastings and Brown (2002a) reported that higher self-efficacy scores predicted lower levels of negative emotional reactions toward learners with challenging behaviors by parents, special educators, and other related service providers. The ability to study such factors as efficacy responses, learning outcomes, and emotional reactions in special educators, parents, and caregivers is possible given earlier work that resulted in suitable measures of domain specific self-efficacy such as the *Difficult Behavior Self-Efficacy Scale* (Hastings & Brown, 2002a). Initially, this scale was designed to assess parents, caregivers, and special education teachers' self-efficacy toward problematic behaviors of individuals with disabilities (Hastings & Brown, 2002a, 2002b); however, this measure has potential for study in the physical education domain following the appropriate study of instrument properties. The availability of suitable measures to study self-efficacy toward individuals displaying challenging behavior in physical education settings is important to facilitate research in this curricular area.

To date only few researchers have examined physical educators' self-efficacy toward challenging learner behavior. This is may be due in part of a lack of suitable instrumentation to measure efficacy responses (Goodwin, 1999). The *Difficult Behavior Self-Efficacy Scale* developed by Hastings and used in Hastings and Brown (2002a) is an instrument that holds promise to further study physical educator

beliefs. Therefore this study was designed to examine the psychometric properties of this scale with a diverse sample of physical educators. A second purpose was to study the factor structure for both the Korean and United States (U.S.) versions of the scale to determine if multigroup invariance was supported and scale properties are stable across cultures. Specifically, this scale was translated to the Korean language (for use with South Korean teachers) and concurrently studied with a U.S. sample to estimate reliability and determine if the hypothesized unidimensional model was supported for the five-item scale.

Methods

Participants

Participants were in-service physical educators from Korea and the U.S., including general physical educators, adapted physical educators, and dual health-physical education teachers from two annual summer workshops. Usable data were collected from 206 Korean physical educators who attended the Physical Educator Regular Education Conference in Seoul, Korea. For the Korean sample, an additional 23 teachers responding to mailed questionnaires were also included in the data analyses. Mail survey responses were secured from both Korean and U.S. conference lists to go along with conference attendees to boost sample size without increasing sample heterogeneity. Based on demographic responses from Korean participants (n = 229), the sample included 33 female and 196 male physical educators (M = 32.91 years, SD = 5.14; age range = 45 years). Other demographic information included the following: 97% of participants were identified as general physical education teacher-coaches, and 3% adapted physical educators. Of these participants, 98% of participants worked at junior high-high school levels and the remaining 2% indicated other settings.

Physical educators from the Midwestern U.S. (n = 139) included 104 females and 35 males (M = 43.21 years, SD = 8.94; age range = 47 years). Demographic information included the following: 42% of participants identified as general physical education teacher-coaches, 27% were physical-health educators, and the remaining participants indicated other or adapted physical education positions. Fifty percent of participants worked at elementary schools, 38% of participants worked at junior high-high schools, and the remaining 12% indicated other workplaces. Sampling methods for the U.S. participants resulted from two potential pools of participants, including current participants of the Physical Education Summer Institute workshop (n = 67) and past participants of this same conference who were not present at the time of initial data collection (n = 72). Those not present were surveyed using mailed packets sent out after the workshop consistent with data collected on Korean participants who did not show up for their conference. This study was approved by the Institutional Review Board and included use of a study information sheet that was provided to all participants before their taking part in the study.

Measures

The *Difficult Behavior Self-Efficacy Scale*, developed by Hastings and Brown (2002a), was used to assess participants' self-efficacy toward challenging behav-

ior when working with children with autism. The *Difficult Behavior Self-Efficacy Scale* was adopted from measure of self-efficacy in staff, caregivers, and parents (Hastings & Brown, 2002a; Hastings & Symes, 2002). This scale has five items that addressed self-efficacy dimensions, including feelings of confidence, personal difficulty dealing with challenging behaviors, dealing with challenging behaviors in a positive way, satisfaction in dealing with children with challenging behaviors, and feelings of control of challenging behaviors. Specific items included statements about dealing with challenging behaviors in relation to "how confident," "how difficult," "positive effect," "satisfaction," and "control" personal beliefs held by respondents. The Likert-type response selections ranged from 1 to 7 (see Figure 1). Responses were then summed for the five items resulting in a total self-efficacy score. Existing alpha reliability estimates for this scaled include high internal consistency scores ($\alpha = 0.89$) reported in Hastings and Brown (2002a); however, there is no known validity estimates for the *Difficulty Behavior Self-Efficacy Scale*.

A brief video clip that shows the example of aggressive challenging behavior was provided to respondents before filling out the five item instrument. For the U.S. sample, an American teacher and an adolescent were engaged in an incident

(SE 1) How confident are you in dealing with challenging behaviours of the child/children with autism you care for?							
	1	2	3	4	5	6	7
	Not at all Confident						Very Confident
(SE 2) How difficult do you personally find it to deal with challenging behaviours of the child/children with autism you care for?							
	1	2	3	4	5	6	7
	Very Difficult						Not at all difficult
(SE 3) To what extent do you feel that the way you deal with the challenging behaviours of the child/children with autism you care for has a positive effect?							
	1	2	3	4	5	6	7
	Has no positive effect at all						Has a very positive effect
(SE 4) How satisfied are you with the way in which you deal with the challenging behaviours of the child/children with autism you care for?							
	1	2	3	4	5	6	7
	Not satisfied at all						Very Satisfied
(SE 5) To what extent do you feel in control of the challenging behaviors of the child/children with autism you care for?							
	1	2	3	4	5	6	7
	Not in control at all						Very much in control

Figure 1 — Hastings (2002a) five item self-efficacy scale and response choices.

of challenging behavior to provide respondents with a visual cue helping teachers focus on aggressive challenging behaviors. For the Korean sample, an identical challenging behavior scenario (using a Korean teacher and child) was created and used to help respondents find a reference point for challenging learner behavior. Both videos contained a male student dribbling a basketball. The student was identified to participants as a child with autism spectrum disorder. When the teacher introduced game instructions, the student demonstrated the challenging behaviors including refusal of the teacher's request, foul language, and aggressive behavior (threw the basketball in the direction of the teacher). Each video clip lasted 1 min and the panel of experts who were involved with translating the questionnaire reviewed content validity of the video clips. They viewed and discussed the comparable language and student's challenging behaviors (i.e., refusal if the teacher's request, foul language, and threw the basketball toward the teacher) in both English and Korean languages. They all agreed that the content for both the Korean and U.S. versions of the video clip represented similar challenging behavior episodes. This video stimulus was used to help Korean and U.S. respondents understand the context of aggressive challenging behavior and insured that a physical education specific reference point was established for responses. Both mail and conference respondents were provided video stimulus to view before filling out the scale.

Translation Procedures

The *Difficult Behavior Self-Efficacy Scale* was translated into Korean and pilot study was conducted to estimate reliability for Korean respondents. Translation of the English version of the scale into Korean included steps suggested by Banville, Desrosiers, and Genet-Volet (2000). The translation was supported by a panel of four bilingual adapted physical education professionals. First, back translation was performed by a panel of experts that included two doctoral students and two professors. Two doctoral students who were studying adapted physical education in the U.S. translated the original version into Korean, and two bilingual professors who were working in adapted physical education in the U.S. back translated into English. Second, the same four bilingual professionals, who worked in the first step, compared the original version to the back translated version. Third, the same panel of experts in the previous steps determined content validity. According to suggestions from the panel of experts, content and grammar edits were made and then the instrument was resubmitted to the panel for their final review. Fourth, to get a reliability estimate of the translated Korean version, an initial pilot study was conducted on 11 Korean respondents before primary data collection. Test-retest reliability estimates were collected via pilot testing using a 2-week interval between responses for pretest and posttest responses. The initial pilot study conducted on 11 separate Korean participants to estimate stability over time of the translated questionnaire resulted in a moderate reliability estimate ($r = 0.64$). The correlation coefficients for the five individual items included 0.60, 0.50, 0.58, 0.75, and 1.00, indicating that the second and third items did not have acceptable test-retest reliability estimates. These values were somewhat low; however, the researchers believed they were adequate to proceed to the primary data collection phase given the low sample size.

Data Collation Procedures

After back translation procedures and initial pilot study, primary data were collected from on site visitations of teacher conferences and mail survey procedures. Participants from the summer conferences (in Korea and the U.S.) watched the 1-min video stimulus at the workshops during the same summer season. In both cases, Korean and U.S. teachers viewed the challenging behavior video clip, which was presented to respondents as "an example" of challenging behavior and completed questionnaires in the same room at their respective conference sites. We also provide a vignette that explains an example of challenging behavior on top of the questionnaire:

> Jeong-Suk (or Patrick for the U.S. participants) is a mail adolescent with autism spectrum disorder. Sometimes Jeong-Suk is defiant and even aggressive toward the teacher and classmates during the physical education class. He will refuse to comply with simple request and has been known to be physically aggressive toward peers, teachers, and even his parents. Respondents were instructed and monitored to insure individual responses to items by discouraging discussion during data collection.

To secure an adequate sample, additional mailed questionnaires were sent out to both Korean and U.S. physical educators. The mailing procedures followed Salant and Dillman's (1994) survey data collection methods. Salant and Dillman recommended a prestudy post card outlining intent to send respondents a packet in the near future. A mailed packet followed this initial post card and included cover letter, self-addressed envelope, questionnaire, compact disk (containing video stimulus), human subject participant information sheet, and a token. Along with the mailed packets, participants were provided a cover letter that specified directions for respondents indicating the order of viewing the video clip before filling out the questionnaire. This specifically indicated that the questionnaire was in reference to a range of challenging behavior with the video serving as an example for children with autism. Two weeks after this initial mailing, a postcard reminder followed with a request to return the original questionnaire. After an additional two weeks, a second mailed packet was sent and this was again followed up by a postcard reminder two weeks later.

Data Analyses

The PRELIS and LISREL programs (Jöreskog & Sörbom, 1993) were used to compute the descriptive statistics and the confirmatory factor analysis for the total sample and for each of the two subsamples (Korean and U.S.). Tests of univariate and multivariate normality (Mardia, 1970) were used to examine the skewness and kurtosis levels in the sample data. For the confirmatory factor analysis, the maximum likelihood factor extraction was used with a variance-covariance (CM) matrix according to the recommendation of Olsson, Foss, Troye, and Howell (2000). To study if the inferences by Hastings and Brown (2002a) concerning the unidimensional nature of the self-efficacy scale, a one-factor model was tested with confirmatory factor analysis. Specifically, a one-factor test was made to determine whether a single factor would be confirmed for the 5-item scale.

Various fit indices were used to test the fit of the one-factor model. The Standardized Root Mean Square Residual (SRMR; Bentler, 1995) and the Goodness of Fit (GFI; Jöreskog & Sörbom, 1984) indices were used as absolute indices. The SRMR should be below 0.08 to represent good model fit (Hu & Bentler, 1999). The GFI should be above 0.95 for good model fit. The Root Mean Square Error of Approximation (RMSEA; Steiger, 1990; Steiger & Lind, 1980) was used and Hu and Bentler (1999) recommended that it should be below 0.06 for good model fit. The Comparative Fit Index (CFI; Bentler, 1990) and the Nonnormed Fit Index (NNFI; Tucker & Lewis, 1973) were also used; Hu and Bentler (1999) indicated these should be above 0.90 for good model fit.

To further study the unidimensional model and determine estimates of invariance between Korean and U.S. samples, structural equation modeling analyses were run using LISREL 8.7. The invariance procedures of Meredith (1993) and Vandenberg (2002) were used to test whether (a) the one-factor model was a good representation for the data from Korea and U.S. (confirgural invariance), (b) the factor loadings were invariant across the two samples (metric invariance), (c) the variances for the two groups were invariant, (d) the means of the five items were invariant across the two samples (scalar invariance), and (e) the measurement errors were invariant across the two samples.

Internal consistency of the five items of the one-factor self-efficacy scale was estimated by using the weighted omega coefficient as suggested by Bacon, Sauer, and Young (1995). The average variance extracted (AVE) was estimated with procedures outlined by Fornell and Larcker (1981). The AVE depicts the amount of variance explained by the one factor and should be above 0.50.

Results

The Korean and U.S. returning samples for the primary study included 100% of the conference attendees from both Korean and U.S. summer workshops. Moreover, a total of 106 questionnaires returned (46%) from the mail survey portion of the data resulted in 368 usable data sets. The results are presented in the following sections: (a) Descriptive Statistics, (b) Confirmatory Factor Analysis, (c) Internal Consistency Reliability and Variance Extracted, and (d) Invariance Testing.

Descriptive Statistics

Table 1 provides item statistics that indicate the range of scores for the sample; separate analyses are included for the Total sample ($N = 368$), the Korean sample ($n = 229$), and the U.S. sample ($n = 139$). Skewness and Kurtosis values, found in Table 1, with univariate and multivariate tests (Mardia, 1970) are also included. Means and standard deviations for both Korean and U.S. samples demonstrate scores that reflect middle or undecided responses to self-efficacy statements.

For the total sample (Both Korean and U.S. samples), three of the items were significantly ($p < .05$) negatively skewed indicating more respondents scored high on the scale. For the total sample, four of the items had significant ($p < .05$) negative kurtosis indicating the curves were flatter than a normal curve or were platykurtic. None of the item scores for the Korean sample were significantly ($p > .05$) different

Table 1 Descriptive Statistics for Five Self-Efficacy Items for Total Sample (N = 368), Korean Sample (n = 229), and United States Sample (n = 139)

Item	Mean	*sd*	Skewness	*z*	*p*	Kurtosis	*z*	*p*
Total Sample								
SE1	4.33	1.44	– .28	– 2.06	0.04	– .62	– 3.36	0.00
SE2	3.65	1.41	– .08	– .86	0.38	– .72	– 4.32	0.00
SE3	4.57	1.45	– .28	– 2.06	0.04	– .54	– 2.72	0.01
SE4	4.12	1.37	– .33	– 2.25	0.02	– .26	– .98	0.32
SE5	4.18	1.32	– .23	– 1.87	0.06	– .52	– 2.62	0.01
Multivariate Tests								
Skewness z = 5.64, p < .05; Kurtosis z = 4.98, p < .05								
Skewness and Kurtosis Chi Square = 56.69, p < .05								
Korean Sample								
SE1	3.96	1.40	0.00	0.03	0.97	– .65	– 2.82	0.01
SE2	3.36	1.42	0.12	1.03	0.31	– .77	– 3.65	0.00
SE3	4.34	1.49	– .09	– .82	0.41	– .74	– 3.46	0.00
SE4	3.79	1.38	– .24	– 1.68	0.09	– .40	– 1.36	0.19
SE5	3.97	1.30	– .21	– 1.58	0.12	– .57	– 2.26	0.02
Multivariate Tests								
Skewness z = 4.65, p < .05; Kurtosis z = 3.58, p < .05								
Skewness and Kurtosis Chi Square = 34.48, p < .05								
United States Sample								
SE1	4.94	1.27	– .83	– 4.03	0.00	0.68	1.61	0.11
SE2	4.12	1.24	– .26	– 1.25	0.21	– .35	– .77	0.44
SE3	4.94	1.30	– .55	–- 2.68	0.01	0.28	0.90	0.37
SE4	4.66	1.18	– .32	– 1.57	0.12	– .11	– .02	0.99
SE5	4.52	1.27	– .28	– 1.35	0.18	– .51	– 1.37	0.17
Multivariate Tests								
Skewness z = 4.53, p < .05; Kurtosis z = 3.26, p < .05								
Skewness and Kurtosis Chi Square = 31.25, p < .05								

Note. SE1 to SE5 are the five self-efficacy items.

from zero skewness; however, four of the items displayed platykurtic curves. For the U.S. sample, two of the items were significantly (p < .05) negatively skewed; yet, none of the items differed significantly (p > .05) from mesokurtic form. All of the multivariate normality tests were significant (p < .05), indicating these data do not display multivariate normality. Olsson et al. (2000), however, indicated that a maximum likelihood estimation method would be more appropriate with data similar to this with less than 2,000 participants.

Confirmatory Factor Analyses

The results of the confirmatory factor analyses (CFA) computed separately for the Total, Korean, and U.S. samples are presented in Table 2. The chi square statistic for each of the three analyses was significant ($p < .05$); this is often the case when analyzing relatively large numbers of respondents. Model fit is normally tested with a variety of fit statistics. The one-factor model was tested in each case and this resulted in good model fit for each of the samples when considering SRMR (0.03 less than 0.08), GFI (0.97 greater than 0.95), CFI (0.98 greater than 0.90), and NNFI (0.95 greater than 0.90). The only fit statistic which indicated less close fit was RMSEA (0.12, 0.11, and 0.10 for the three analyses were not less than the expected 0.06). Overall, acceptable model fit was found for the one-factor model of self-efficacy for the five-item scale.

The completely standardized factor loadings for the total sample ranged from 0.64 to 0.82, from 0.61 to 0.83 for the Korean sample, and from 0.60 to 0.83 for the U.S. sample. The measurement errors (item theta delta values) ranged from 0.32 to 0.59 for the total sample, from 0.31 to 0.63 for the Korean sample, and from 0.30 to 0.64 for the U.S. sample. The second item consistently resulted in the weakest factor fit. The completely standardized parameter estimates for the total sample are depicted in Figure 2; similar values were obtained for each of the two subsamples.

Internal Consistency Reliability and Variance Extracted

The weighted omega, internal consistency reliability values for the one factor were 0.88 for all three analyses including total sample, Korean sample, and U.S. sample. These represent high internal consistency for the five items of the scale. The AVE value was 0.59 for the total sample, 0.57 for the Korean sample, and 0.57 for the U.S. sample. The one factor explained more than 50% of the total variability with all AVE values greater than 0.50.

Table 2 Confirmatory Factor Analysis for the Total Sample, Korean Sample, and United States Sample

Sample	χ^2	*df*	RMSEA	SRMR	GFI	CFI	NNFI	CN
Total Sample ($n = 368$)	27.44[a]	5	0.12	0.03	0.97	0.98	0.95	202.84
Korean Sample ($n = 229$)	18.58[a]	5	0.11	0.03	0.97	0.97	0.95	186.14
U.S. Sample ($n = 139$)	12.40[a]	5	0.10	0.03	0.97	0.98	0.95	168.97

[a]Chi Square significant ($p < .05$)

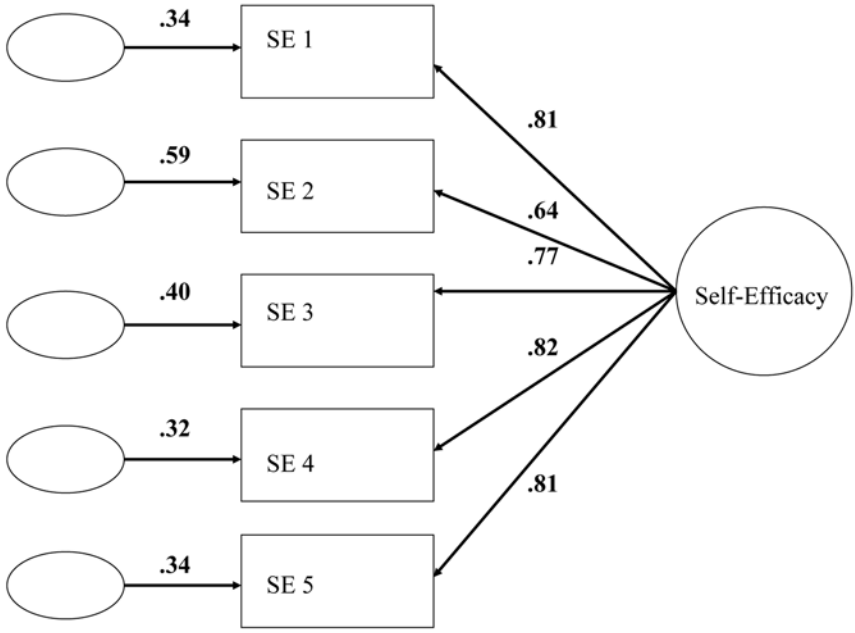

Figure 2 — Path analysis for structure of latent variable using standardized values for the entire sample ($N = 368$).

Invariance Testing

The results of the invariance testing are summarized in Table 3. The steps to invariance testing vary across authorities; the order for testing used a combination of procedures used by Meredith (1993) and Vandenberg (2002). The invariance testing determined if the parameter estimates for the two subsamples, Korean and U.S. groups, were similar; to have a scale that is applicable to various groups of respondents, the parameter estimates for the model should be invariant across samples.

The first step is to test a baseline model to determine whether the one-factor model is applicable to both samples. Before this the model needs to fit the separate samples adequately; this was established by the separate CFAs for each of the samples in the previous analysis. In this baseline test, all parameters (factor loadings, variances of the latent variable, and measurement errors) were allowed to be freely estimated separately for each of the two samples. The variance-covariance matrices for each of the samples were entered into a stacked program. Adequate fit of this baseline model is necessary to continue with the analysis. The fit statistics

Table 3 Invariance Testing Comparing the Korean and United States *Parameter Estimates*

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	RMSEA	SRMR	NNFI	CFI	CN
Baseline Model	30.97	10			0.11	0.03	0.95	0.97	275.22
Model 1 Lx Constrained	31.31	14	0.34[a] (M1-B)	4	0.08	0.03	0.97	0.98	341.63
Model 2 Tau Constrained	48.11	18	16.80[b] (M2-M1)	4	0.09	0.04	0.96	0.96	265.74
Model 3 Lx & Phi Constrained	32.74	15	1.43[c] (M3-M1)	1	0.08	0.08	0.97	0.97	342.77

Note: B = Baseline Model

M1 = Model 1 = Lx constrained

M2 = Model 2 = Tau coefficients constrained

M3 = Model 3 = Lx and Phi matrices constrained

[a]Table Chi Square ($df = 4$) = 9.48; $p > .05$

[b]Table Chi Square ($df = 4$) = 9.48; $p < .05$

[c]Table Chi Square ($df = 1$) = 3.84; $p > .05$

for this baseline model were adequate; the only fit statistic that indicated less than adequate fit was RMSEA. The 0.90 confidence interval for RMSEA was 0.06–0.15. The CN of 275.22 was an indication of adequate power for making the test of model fit. The one-factor model appeared to be an adequate representation of the data for both subgroups which would support the configural validity (Vandenberg, 2002) of the model across groups.

The second step (Model 1 in Table 3) was a test of whether the factor loadings or Lambda X (Lx) values were invariant across groups. In this program, the Lx values were constrained to be equal for the two samples. All other parameter estimates were allowed to be freely estimated for the two groups. Use of the chi square difference test (Bentler & Bonett, 1980) to compare the chi square value for this Model 1 with the Baseline model chi square, allows one to determine whether the Lx values are invariant. The change in chi square was 0.34, and the change in degrees of freedom was 4; thus, the difference in the chi square values for the Model 1 and Baseline was not significant ($p > 0.05$). Without a significant change in chi square, one may conclude that holding the Lx values equal for the two samples did not make a difference in model fit; consequently, the factor loadings are deemed invariant across the two samples. Vandenberg (2002) labeled this as the scale having metric invariance across groups.

Meredith (1993) and Vandenberg (2002) suggested that the scalar invariance of the scale be tested next. This was the Model 2 test in Table 3 whereby the means

of the five indicator variables (tau coefficients) are now constrained (leaving the Lx values still constrained as well), and the chi square difference comparing Model 2 minus Model 1 chi square values are tested. The chi square difference, with a difference in degrees of freedom of 4, was 16.80 which was significant ($p < .05$), indicating that the means of the indicators for the Korean and U.S. samples were not similar in some manner. The test of differences in the latent variable means for the two samples is found in the kappa test; the mean difference in latent variable means was 0.82, Standard Error of 0.12, and *t*-ratio of 6.43. Thus, the latent variable mean for the U.S. sample was significantly higher than the mean for the Korean sample. To study these mean differences more carefully, independent groups' *t*-ratios were computed to compare the means for the two samples for each of the five indicator variables (Table 4). For all five of the indicator variables, the means for the U.S. sample were significantly ($p < .05$) higher compared with the Korean sample. Model 3 was used to determine if the scale had equal variance estimates for the latent variable; only the variance of the one latent variable would be in the phi matrix in this one-factor model. The chi square difference test was not significant ($p > .05$), which was an indication that the variances were invariant across the two samples. The test of invariance of the measurement errors (theta delta values) was deemed inappropriate because of a lack of scalar invariance.

Table 4 Independent Groups t-Ratios Comparing Mean Scores on Five Indicator Variables on Self-Efficacy Scale

Item	Means	Mean Diff	*SE*	*t*	*p*
SE1					
Korean	3.96	– .98	0.14	-6.90	0.00
U.S.	4.94				
SE2					
Korean	3.36	– .76	0.14	–5.41	0.00
U.S.	4.12				
SE3					
Korean	4.34	– .60	0.15	–4.06	0.00
U.S.	4.94				
SE4					
Korean	3.79	– .86	.15	–6.15	0.00
U.S.	4.65				
SE5					
Korean	3.97	–.54	0.14	–3.91	0.00
U.S.	4.52				

Discussion

The *Difficult Behavior Self-Efficacy Scale* developed by Hastings and Brown (2002a) has been used to estimate self-efficacy in special educators, caregivers, and now physical educators. The current study is the first attempt to estimate the factorial validity and internal consistency reliability of the scale with the responses from physical educators to issues of challenging learner behavior. The invariance of the scale across two samples, a Korean and U.S. group of physical educators, was tested. From CFAs on the total sample, Korean sample, and U.S. sample, the factorial validity of the one-factor model for the scale was supported with adequate model fit in each situation. The internal consistency of the scale was high for the total group and for the subgroups of respondents. In invariance testing, the scale demonstrated configural validity (one-factor model) across the Korean and U.S. samples; furthermore, metric invariance was also confirmed with factor loadings and variances of the latent variable invariant across the two samples. Scalar invariance was not supported; the five items on the scale all had means that were significantly higher for the U.S. sample compared with the Korean sample. Further work needs to be done to insure that items are invariant across various groups so that a scale exists for measuring different groups on the same scalar basis.

These data support recommendations to use caution when attempting to make cross cultural comparisons with instruments translated from one language to another and attempting to study complex concepts and ideas (Banville et al., 2000). Disability as an educational concept and issue for teachers to reflect on is very complex, particularly in relation to challenging learner behavior (Hastings & Brown, 2002a). The rigorous methodology used in this study is consistent with recommendations by Banville et al. (2000); however, more study is needed. Although the concepts studied demonstrate common themes related to behavioral disabilities and teacher efficacy that translate from English to Korean languages, results point to the need to consider additional differences across samples. These include the potential for variability within the construct across cultures.

Our efforts resulted in support for measuring the latent construct but clear variability in responses between group means. This may point to a need to consider cultural influence and the potential for participants from different backgrounds to differ uniformly in perceptions of aggression depicted in the video stimulation or wording in relation to efficacy statements. The point is not that one group has higher or lower scores but that differences may exist between groups that explain the lack of scalar invariance. The different means found in U.S. and Korean samples make it likely that study of teacher behaviors across samples would have to consider the differences within the construct to make valid comparisons. Further, Kozub and Lienert (2003) point out that the importance of teacher attitudes, perceptions, or beliefs are rooted in how these variables predict teacher behavior. These include self-efficacy as a predictor of important teacher behavior, such as a willingness to develop teaching skills and use of teaching innovations (Meijer & Foster, 1988). The lack of scalar invariance does not preclude study of teacher behaviors as a function of self-efficacy using the current measure. Our results indicate a need to use caution when comparing findings in a cross cultural setting using English and Korean samples, however.

In conclusion, the one-factor model for the *Difficult Behavior Self-Efficacy Scale* developed by Hastings and Brown (2002a) has factorial validity and reliability when used to study these physical educators from Korea and the U.S. Invariance testing with subgroups of teachers from Korea and U.S. revealed the five-item scale has configural and metric invariance for the two groups; however, the items lacked scalar invariance across groups.

Limitations of Study

A few limitation of the current study should be pointed out. The use of a single challenging behavior example video clip may limit the generalization of participants' perceptions on all challenging behavior. One of the rationales for the use of a video clip was to create a comparable context for responses irrespective of teacher exposure to children with ASD. The specific nature of the video and vignette to a child with ASD is a limiting factor in the current study and points to a need for continued study on other disability categories, a wider range of challenging behaviors in actual teaching settings. Another limitation of the current study is the small number of U.S. participants as well as differences in proportions of females to males between the two samples. This along with potential differences in level of teaching where Korean teachers were primarily from secondary grades and U.S. participants where overrepresented in relation to elementary physical educators is an issue that needs attention in future sampling plans. Our responding sample did not afford an identical match in relation to teaching level or gender. Finally, the measure used in the current study is an adequate scale to access physical education teachers' self-efficacy toward challenging behaviors; however, this scale is not completely based on Bandura's self-efficacy theory and recommendations for instrument development. Only three items (i.e., item 1, 2, and 5) from the current scale are in line with Bandura's self-efficacy conceptualization (Figure 1). The remaining two items reflect if teaching methods used by physical educators for dealing with challenging behaviors have positive effects (item 3) and the extent that teacher satisfaction resulted in dealing with challenging behaviors (item 4). These two items (item 3 and 4) may confound study of self-efficacy with respect to application of Bandura's self-efficacy model. More study is needed to determine how issues related to teacher perceptions of outcomes and satisfaction relate to the construct of self-efficacy.

Recommendations for Future Research

Future researchers should focus on continued instrument development and additional exploration of self-efficacy toward challenging behavior in physical educators teaching children with disabilities. Further, studies of how groups who vary in training within and across cultures are needed. It is likely that the nature of the current samples, volunteers attending national conferences, are a select group within each culture and broader samples including teachers not attending workshops are needed to create an externally valid measure of self-efficacy. Teacher education curriculum should impact self-efficacy responses. The more knowledge educators have about aggressive behavior and related perceptions of self efficacy need to be

studied as predictors of important teacher behaviors. Finally, gender effects and the potential invariance between males and females are in need of study with a sample that contains equal proportions of male and female physical educators.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

Bacon, D.R., Sauer, P.L., & Young, M. (1995). Composite reliability in structural equations modeling. *Educational and Psychological Measurement, 55,* 394–406.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman and Company.

Banville, D., Desrosiers, P., & Genet-Volet, Y. (2000). Translating questionnaires and inventories using a cross-cultural translation technique. *Journal of Teaching in Physical Education, 19,* 374–387.

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

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

Bentler, P.M., & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structure. *Psychological Bulletin, 88,* 588–606.

Brouwers, A., & Tomic, W. (2001). The factorial validity of scores on the teacher interpersonal self-efficacy scale. *Educational and Psychological Measurement, 61,* 433–445.

Deemer, S.A., & Minke, K.M. (1999). An investigation of the factor structure of the teacher efficacy scale. *The Journal of Educational Research, 93,* 3–10.

Feltz, D. L (2000). Self-efficacy beliefs of athletes, teams, and coaches. *Korean Association of Health, Physical Education, and Dance, August,* 24-27.

Fornell, C., & Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *JMR, Journal of Marketing Research, 18,* 39–50.

Goodwin, L.D. (1999). The role of factor analysis in the estimation of construct validity. *Measurement in Physical Education and Exercise Science, 3,* 85–100.

Hastings, R.P., & Brown, T. (2002a). Behavioural knowledge, causal beliefs, and self-efficacy as predictors of special educators' emotional reactions to challenging behaviours. *Journal of Intellectual Disability Research, 46,* 144–150.

Hastings, R.P., & Brown, T. (2002b). Behavior problems of children with autism, parental self-efficacy, and mental health. *American Journal of Mental Retardation, 107,* 222–232.

Hastings, R.P., & Symes, M.D. (2002). Early intensive behavioral intervention for children with autism: Parental therapeutic self-efficacy. *Research in Developmental Disabilities, 23,* 332–341.

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

Jöreskog, K.G., & Sörbom, D. (1984). *LISREL VI user's guide* (3rd ed.). Mooresville, IN: Scientific Software.

Jöreskog, K.G., & Sörbom, D. (1993). *LISREL 8: User's reference guide*. Chicago: Scientific Software International.

Kaiser, B., & Rasminsky, J.S. (2003). *Challenging behavior in young children: Understanding, preventing, and responding effectively*. Boston, MA: Allyn and Bacon.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2,* 217–253.

Kozub, F.M., & Lienert, C. (2003). Attitudes towards teaching children with disabilities: A review of literature and proposed paradigm for future research. *Adapted Physical Activity Quarterly, 20,* 323–346.

LaMaster, K., Gall, K., Kinchin, G., & Siedentop, D. (1998). Inclusion practices of effective elementary specialists. *Adapted Physical Activity Quarterly, 15,* 64–81.

Maddux, J.E. (1995). *Self-efficacy, adaptation, and adjustment: Theory, research, and application*. New York: Plenum Press.

Malete, L., & Feltz, D.L. (2000). The effect of a coaching education program on coaching efficacy. *The Sport Psychologist, 14,* 410–417.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*(3), 519–530.

Martin, J.J., & Hodges-Kulinna, P. (2004). Self-efficacy theory and the theory of planned behavior: Teaching physically active physical education classes. *Research Quarterly for Exercise and Sport, 75,* 288–297.

Martin, J.J., & Kulinna, P.H. (2003). The development of a physical education teachers' physical activity self-efficacy instrument. *Journal of Teaching in Physical Education, 22,* 219–232.

Martin, J.J., & Kulinna, P.H. (2005). A social cognitive perspective of physical activity related behavior in physical education. *Journal of Teaching in Physical Education, 24,* 265–281.

Martin, J.J., Kulinna, P.H., Eklund, R.C., & Reed, B. (2001). Determinants of teachers' intentions to teach physically active physical education classes. *Journal of Teaching in Physical Education, 20,* 129–143.

Martin, J.J., McCaughtry, N., Kulinna, P., Cothran, D., & Faust, R. (2008). The effectiveness of a mentoring-based professional development on physical education teachers' pedometer and computer efficacy and anxiety. *Journal of Teaching in Physical Education, 27*(1), 68–82.

Meijer, C.J.W., & Foster, S.F. (1988). The effect of teacher self-efficacy on referral chance. *The Journal of Special Education, 22,* 378–385.

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

Olsson, U.H., Foss, T., Troye, S.V., & Howell, R.D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling, 7,* 557–595.

Ross, J.A., Cousins, J.B., & Gadalla, T. (1996). Within-Teacher Predictors of Teacher Self-Efficacy. *Teaching and Teacher Education, 12,* 385–400.

Salant, P., & Dillman, D.A. (1994). *How to conduct your own survey*. New York: John Wiley & Sons.

Steiger, J.H. (1990). Structural model evaluation: An interval estimation approach. *Multivariate Behavioral Research, 25,* 173–180.

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

Tucker, L.R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1–10.

Vandenberg, R.J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Occupational Research Methods, 5,* 139–158.

Wheatley, K.F. (2002). The potential benefits of teacher efficacy doubts for educational reform. *Teaching and Teacher Education, 18,* 5–22.