ACADEMIC SELF-EFFICACY AMONG JUNIOR HIGH SCHOOL STUDENTS IN GHANA: EVALUATING FACTOR STRUCTURE AND MEASUREMENT INVARIANCE ACROSS GENDER

DAVID ANSONG, SARAH R. EISENSMITH, RAINIER D. MASA, AND GINA A. CHOWA

University of North Carolina at Chapel Hill

Self-efficacy is a universal construct, but few validated measures exist for researchers in developing countries to use in assessing youths' perceptions of their ability to achieve academic success. This study examined the cross-cultural suitability and psychometric properties of an academic self-efficacy scale (ASES) adapted for the Ghanaian context. ASES construct validity was assessed with a sample of 4,289 Ghanaian junior high-school students and exploratory and confirmatory factor analysis. Invariance testing assessed the scale's measurement equivalence by gender and temporal stability of gender equivalence. The ASES is a valid, reliable one-dimensional scale for assessing young Ghanaians' perceptions of their academic capabilities, and it works equally well across genders. As adapted, ASES is a valid scale with utility for researchers examining predictors and effects of academic self-efficacy. The ASES has important implications for decisions regarding investment in programs aimed at improving academic self-efficacy of youth, both in sub-Saharan Africa and the increasingly diverse American public schools. © 2016 Wiley Periodicals, Inc.

Youths' academic self-efficacy is considered fundamental to academic success and overall well-being. Researchers who study the developmental effects of asset ownership have suggested that self-efficacy plays an important mediating role in the relationship between economic resources and educational outcomes (Chowa, Masa, Anson, & Ramos, 2015; Elliott, Sherraden, Johnson, Johnson, & Peterson, 2007). Conceptually, students' concerns about having sufficient financial resources to pay for education-related expenses might lead to self-doubts about their academic abilities, which in turn, might interrupt learning efforts and diminish academic performance (Chowa et al., 2015).

Despite the relevance of academic self-efficacy as a potential determinant of academic performance of students experiencing economic stress, measurement of the academic self-efficacy construct has received little research attention in developing countries. This research gap persists, at least in part, because the vast majority of emerging research on self-efficacy in developing countries has focused on general self-efficacy measures (Yendork & Somhlaba, 2015) or self-efficacy in the health domain (Asante & Doku, 2010; Baah-Odoom, & Riley, 2013). The few studies that have examined academic self-efficacy in sub-Saharan Africa have focused on older youth in higher education institutions (Gota, 2012; Matoti, 2011).

This study sought to close this research gap by focusing on the validation of the Academic Self-Efficacy Subscale (ASES) derived from the Self-Efficacy Questionnaire for Children (SEQ-C; Muris, 2001) and tailored for the Ghanaian context. Measurement scales, such as the SEQ-C, are instrumental to social science research, and the quality of data yielded by these tools is a concern (Cizek, 2012). Adapting measurement instruments is a complex process that requires thought related to content maintenance, psychometric properties, validity for the target population, and cultural fit to the target population, all of which necessitate high methodological rigor (Borsa, Damásio, & Bandeira, 2012). The present study highlights how these best practices were followed to facilitate increased confidence in conclusions drawn from the ASES.

To our knowledge, no prior study has culturally adopted and psychometrically validated the ASES specific to sub-Saharan African youth at the junior high school level. Establishing adequacy

Correspondence to: David Ansong, 325 Pittsboro Street, CB #3550, Chapel Hill, NC 27599. E-mail: ansong @email.unc.edu

of the academic self-efficacy construct in the developing world context is critical to helping education researchers and practitioners to better assess academic self-efficacy of adolescents and offer appropriate interventions to improve academic outcomes. Moreover, this study's test of the construct validity of academic self-efficacy makes an important contribution to ensuring future tests of structural relationships between academic self-efficacy, and its causes and effects are based on sound measurement (e.g., good construct validity). This study also aimed to assess the generalizability of the scale across genders and to determine whether gender equivalence held over time. Such examination is critical if the validated scale is to have wide application in junior high schools because researchers and practitioners must know whether the scale works equally well for both boys and girls over the duration of junior high school.

ACADEMIC SELF-EFFICACY

Self-efficacy refers to an individual's judgment of his or her ability to learn or to achieve a certain level of performance (Bandura, 1986), and it is task and domain specific (Lunenburg, 2011). Thus, different types of self-efficacy beliefs exist that relate to specific domains, including social self-efficacy, emotional self-efficacy, and academic self-efficacy. The current study focused on academic self-efficacy, which relates to the individual's beliefs about his or her ability to achieve self-valued goals or standards in the school context (Muris, 2001). Students' academic self-efficacy is fundamental to their learning because each person's perception of his or her academic ability can influence personal motivation for completing work and how well the student performs in school.

Academic self-efficacy contributes to educational achievement through an increased use of specific strategies and cognitive activities, as well as through the positive impact of efficacy beliefs more generally (Chemers, Hu, & Garcia, 2001). Previous studies have shown that high levels of academic self-efficacy are important to sustaining students' motivation, participating in learning, putting forth effort, achieving desired performance levels, and protecting against academic failure at later stages, as well as other difficulties, such as childhood depression (Bandura, Pastorelli, Barbaranelli, & Caprara, 1999; Britner & Pajares, 2006; Multon, Brown, & Lent, 1991; Schunk, 1985).

Cross-Cultural Differences in Academic Self-Efficacy

Although research has suggested that self-efficacy is a universal construct, some cultural variations exist in how strongly people feel about their abilities (Oettingen, 1997; Scholz, Doña, Sud, & Schwarzer, 2002). In a study of nearly 20,000 participants from 25 countries, Scholz and colleagues (2002) found that notwithstanding the universality of the general self-efficacy construct, there are "a number of cross-cultural differences that merit further investigation" (p. 242). Although the study did not include countries from the African and Australian continents, it lends support to the hypothesis that the intensity of self-efficacy beliefs vary by culture. Contrasted with individualistic-oriented cultures (i.e., the United States and most of northern and western Europe, Australia, and New Zealand), collectivistic cultures (e.g., Africa, Asia, the Middle East, Pacific Islands, and Central and South America) tend to report lower self-efficacy beliefs because of the emphasis on group abilities rather than individual abilities (Schunk & Usher, 2011; Woodward & Denton, 2013; Yan & Gaier, 1994). In collective-oriented cultures, the confidence that students have in their familial and social relations combined with social support from parents were shown to be strong predictors of students' academic performance (Nyarko, 2011; Wu, Tsang, & Ming, 2014).

Although significant research has been conducted on the cross-cultural differences in general self-efficacy, the body of evidence on academic self-efficacy is insufficient. More studies are needed to investigate the psychometric properties of ASES used in non-Western countries, particularly

given the lack of measures of academic self-efficacy that have been validated in developing countries, including Ghana. This study helps fill the knowledge gap by examining the adaptation and appropriateness of a U.S.-based ASES to a non-Western context. Validation of this measure to a non-Western context may also hold relevance for researchers and practitioners in the United States and other developed countries, given the increasing diversity of schools (Joy & Kolb, 2009) and the need for more nuanced and culturally informed measurement tools. Moreover, even though there are over 200 studies on academic self-efficacy in developed countries, literature on the development and psychometric testing of academic self-efficacy for students at the junior high school level is scant.

Gender Differences in Academic Self-Efficacy

Individual differences such as gender can also affect global self-efficacy and perceptions of self-efficacy in specific domains. In the academic domain for instance, research conducted primarily in the United States has suggested that gender variations exist in students' academic self-efficacy regarding certain school subjects, such as science, language arts, or math (Britner & Parajes, 2001, 2006; Bussey, 2011; Else-Quest, Hyde, & Linn, 2010; Usher & Parajes, 2008). Similarly, gender differences in self-esteem among older youth have been reported by a few studies conducted in sub-Saharan Africa (Atindanbila, Winifred, & Awuah-Peasah, 2012; Imhonde, 2013). Gender roles might account for gender differences in academic self-efficacy (Kling, 1999: Nunn & Thomas, 1999). Such differences are particularly possible in non-Western contexts where many families socialize their children by assigning boys to leadership roles that are more likely to build their confidence, whereas they assign girls to domestic roles such as completing household chores.

Given the well-documented gender differences in beliefs, attitudes, and behavior expected of Ghanaian youth, a critical question for this study was whether the same instrument might be equally effective in assessing the academic self-esteem of young Ghanaian boys and girls. Although there is an emerging scientific literature on self-efficacy in Ghana, there are gaps in the development and psychometric testing of academic self-efficacy for junior high school students. This study helps fill the knowledge gap by examining the generalizability of the ASES across gender and the stability of the gender equivalence over time. The goal to examine the temporal stability of gender equivalence addressed an important gap identified in a recent meta-analysis on gender differences in academic self-efficacy (Huang, 2013). A recommendation from Huang's (2013) meta-analysis was the need for future research to "longitudinally examine gender differences in academic self-efficacy to determine the prevalence of gender differences during different life stages" (p. 1).

METHODS

Participants

The current study used baseline and 3-year follow-up data from the YouthSave Ghana Experiment, a large social experiment designed to test the viability of youth savings accounts in Ghana and their impact on educational, psychosocial, health, and financial well-being outcomes of Ghanaian youth. Academic self-efficacy is one of the psychological constructs assessed in the YouthSave Ghana Experiment. The baseline data were collected in 2011 from a sample of 6,252 junior high school pupils who represented eight of the 10 administrative regions of Ghana: Ashanti; Brong Ahafo; Volta; and the Northern, Western, Central, Eastern, and Greater Accra regions. Because one of the goals of the current study was to test the temporal stability of gender invariance, the study's sample was limited to students who completed the academic self-efficacy questions at both baseline and follow-up ($N = 4,289$).

Table 1
Item Pool of Eight Academic Self-Efficacy Questions

Item 1	How well can you get teachers to help you when you get stuck on schoolwork?
Item 2	How well can you study when there are other interesting things to do?
Item 3	How well can you study a chapter for a test?
Item 4	How well do you succeed in finishing all your homework every day?
Item 5	How well can you pay attention during every class?
Item 6	How well do you succeed in passing all subjects?
Item 7	How well do you succeed in satisfying your parents with your schoolwork?
Item 8	How well do you succeed in passing a test?

Procedures

YouthSave researchers used guidelines from the International Test Commission to guide the review and adaptation of the SEQ-C (Muris, 2001) to the Ghanaian context. First, the instrument was reviewed for relevance and fit by a team of social science researchers from three universities (i.e., University of North Carolina at Chapel Hill, University of Ghana, and the Center for Social Development at Washington University in St. Louis) and experts in measurement theory and psychological assessment. Based on the feedback from this expert review, the adapted instrument was further revised and then piloted in two junior high schools in Ghana. Field data were collected using three best-practice methods: interview-administered surveys, cognitive interviews, and in-depth interviews (Beatty & Willis, 2007). The finalized instrument was administered face-to-face by a team of 40 trained interviewers, each of whom held a college degree. This study was approved by the institutional review boards of the University of Ghana, University of North Carolina at Chapel Hill, and Washington University in St. Louis.

Academic Self-Efficacy Scale

The ASES used in the YouthSave study was adapted from a brief questionnaire developed to measure the social, emotional, and academic self-efficacy of children living in the United States (Muris, 2001). The YouthSave ASES consisted of the eight items listed in Table 1. In the original scale, the eight items were scored on a 5-point scale ranging from 1 (*not at all*) to 5 (*very well*). There are disadvantages to treating 5-point response scales as continuous scales and applying normal theory maximum likelihood because of the susceptibility to incorrect standard errors and biased model statistics (Maydeu-Olivares & Joe, 2005). Consistent with prior studies and based on the recommendations of the expert reviewers, the 5-point response format was expanded to an 11-point response scale for each item to improve response variability and closely approximate a continuous scale (Alwin, 1997; Dawes, 2002). Expansion of the scale made it possible to use full information methods when fitting the latent self-efficacy models because it has the theoretical advantage of producing better estimates compared with limited information methods (Joe & Maydeu-Olivares, 2010). Respondents were asked to rate their level of confidence in their ability to manage their learning behavior, master academic subjects, and fulfill academic expectations (Muris, 2001). The revised response options ranged from 0 (*cannot do at all*) to 5 (*moderately can do*) to 10 (*highly certain can do*).

Analysis

To evaluate the factor structure of the adapted YouthSave ASES, we conducted exploratory factor analysis (EFA) with a calibration subsample. Although a prior study (Muris, 2001) had found

a unidimensional factor structure, it was important to examine the factor structure of the scale in the Ghanaian context because the scale was originally developed in the United States and had not been tested with Ghanaian youth. We used the oblique rotation methods for all EFAs. We used confirmatory factor analysis (CFA) with a validation subsample to validate the EFA results. The calibration and validation samples are subsamples randomly generated from the full sample. Notwithstanding the ordinal nature of the response scale, the 11-point response options closely approximate continuous scales; thus, the scale items were analyzed as continuous variables (Babakus, Ferguson, & Jöreskog, 1987; Olsson, 1979). Although the scale items were fairly normally distributed with skewness values less than $\pm$ .5, we used the maximum likelihood estimation with robust standard errors in Mplus 7.11 for all analyses to account for slight violations of nonnormality (Muthén & Muthén, 1998–2012). The maximum likelihood estimation with robust standard errors is robust enough to handle the slight departures from normality (Hau & Marsh, 2004). In addition, given that respondents were nested in schools, we adjusted for clustering at the school level to account for nonindependence of observations. We used the following multiple statistics to evaluate the goodness of all model fits: chi-square statistic (χ^2) and its p-value, root mean square error of approximation (RMSEA; mediocre if .08 to .10, good if < .05), and comparative fit index (CFI; acceptable if > .90, excellent if > .95; Hu & Bentler, 1999; Kline, 2005). Factor loadings equal to or greater than .30 were deemed adequate (Costello & Osborne, 2005).

In addition to assessing the fit of the ASES to the Ghanaian context, we examined invariance of the scale across gender. The invariance tests assessed whether the parameters of the academic self-efficacy measurement model were statistically identical for boys and girls at baseline and follow-up. First, we examined separate models for boys and girls to confirm the adequacy of the model for each gender (Bowen & Masa, 2015). Next, we conducted gender invariance tests by estimating a succession of three nested models, starting with the least constrained to the most constrained model: configural, metric, and scalar models (Millsap & Olivera-Aguilar, 2012; Vandenberg & Lance, 2000). We used the chi-square difference test to assess whether the nested comparison models fit the data equally well as the next-least restrictive model (Bryant & Satorrab, 2012). We also assessed the longitudinal stability of the gender invariance results by running identical models at baseline and follow-up.

<center>RESULTS</center>

The sample was almost evenly divided between genders, at 50.15% girls (girls = 2,151; boys = 2,138). All items were treated as continuous variables and were normally distributed. As shown in Table 2, all skewness values (i.e., 0.55 to 1.46) were within acceptable departures from normality (i.e., skewness values of $\pm$ 1.5);

Exploratory Factor Analysis Results

The separate EFAs with the calibration sample at the two measurement occasions produced a one-factor solution. The scree plots also suggest a one-factor solution at the two measurement occasions (Fabrigar, Wegener, MacCallum, & Strahan, 1999). The eigenvalues greater than 1 further confirmed a one-factor solution (i.e., baseline: 3.26; follow-up: 3.59). As shown in Table 3, the item loadings at baseline (.38 to .76) and follow-up (.40 to .79) points are all greater than the .30 cutoff (Costello & Osborne, 2005). As shown in Table 4, the one-factor EFA model had an acceptable fit to the data at baseline (χ^2 = 114.65, df = 20, p < .001, RMSEA = .047, 90% CI [.039, .055], CFI = .95) and follow-up (χ^2 = 156.61, df = 20, p < .001, RMSEA = .056, 90% CI [.048, .065], CFI = .95).

Table 2
Summary Statistics of the Observed Academic Self-Efficacy Items at Two Measurement Occasions

| | Full Sample (N = 4,289) | | | | | | Boys (n = 2,138) | | | | | | Girls (n = 2,151) | | | | | |
| | Baseline | | | Follow-Up | | | Baseline | | | Follow-Up | | | Baseline | | | Follow-Up | | |
	M	SD	Skewness	M	SD	Skewness	M	SD	Skewness	M	SD	Skewness	M	SD	Skewness	M	SD	Skewness
Item 1	6.35	3.21	−0.65	6.21	3.16	−0.54	6.43	3.19	−0.67	6.24	3.17	−0.57	6.26	3.23	−0.63	6.18	3.15	−0.51
Item 2	6.23	3.04	−0.64	5.97	2.96	−0.41	6.19	3.11	−0.62	5.99	2.99	−0.43	6.27	2.96	−0.65	5.94	2.94	−0.38
Item 3	8.02	1.88	−1.06	7.61	2.10	−0.78	8.06	1.91	−1.19	7.67	2.08	−0.84	7.98	1.86	−0.92	7.55	2.11	−0.71
Item 4	8.58	1.72	−1.29	8.05	1.98	−1.07	8.61	1.76	−1.46	8.04	2.03	−1.13	8.55	1.68	−1.09	8.07	1.94	−0.98
Item 5	8.63	1.65	−1.41	8.39	1.76	−1.21	8.62	1.67	−1.46	8.42	1.76	−1.32	8.63	1.62	−1.35	8.37	1.76	−1.09
Item 6	7.44	1.88	−0.55	7.48	1.89	−0.73	7.48	1.93	−0.63	7.53	1.91	−0.88	7.41	1.83	−0.46	7.44	1.88	−0.57
Item 7	8.03	1.91	−1.07	7.91	1.98	−1.06	8.09	1.91	−1.12	7.97	1.97	−1.14	7.97	1.91	−1.02	7.86	1.99	−0.99
Item 8	7.89	1.76	−0.76	7.82	1.77	−0.82	7.96	1.78	−0.84	7.89	1.74	−0.84	7.82	1.73	−0.68	7.74	1.80	−0.79

Note. Baseline = first measurement occasion; follow-up = second measurement occasion; *SD* = standard deviation.

Table 3
Factor Loadings of Tested Models

	Baseline				Follow-Up			
	EFA with Calibration Sample	CFAs with Validation Sample			EFA with Calibration Sample	CFAs with Validation Sample		
	Sample	All Groups	Boys Only	Girls Only	Sample	All Groups	Boys Only	Girls Only
Item 1	0.394	0.401	0.387	0.393	0.401	0.406	0.364	0.331
Item 2	0.388	0.400	0.384	0.384	0.413	0.423	0.471	0.417
Item 3	0.592	0.607	0.602	0.588	0.602	0.620	0.663	0.664
Item 4	0.482	0.502	0.525	0.536	0.571	0.587	0.621	0.570
Item 5	0.497	0.516	0.533	0.463	0.637	0.651	0.700	0.615
Item 6	0.694	0.647	0.622	0.602	0.688	0.644	0.652	0.638
Item 7	0.668	0.672	0.653	0.636	0.691	0.691	0.723	0.724
Item 8	0.756	0.715	0.748	0.685	0.794	0.760	0.744	0.746

Note. CFA = confirmatory factor analysis; EFA = exploratory factor analysis; baseline = first measurement occasion; follow-up = second measurement occasion. All loadings are statistically significant at the .001 significance level.

Table 4
Goodness-of-Fit Indices for Academic Self-Efficacy Scale at Baseline and Follow-Up

Models	$\chi^2(df)$***	RMSEA [90% CI]	CFI	Loadings
Baseline				
EFA with calibration sample ($n = 2{,}145$)	114.65(20)	.047 [.039, .055]	.954	.38 to .76
CFA with validation sample ($n = 2{,}145$)	114.65(19)	.047 [.039, .055]	.954	.39 to .76
CFA with validation sample (with correlated errors; $n = 2{,}145$)	93.93(19)	.043 [.034, .052]	.964	.40 to .72
Follow-Up				
EFA with calibration sample ($n = 2{,}145$)	156.61(20)	.056 [.048, .065]	.095	.40 to .79
CFA with validation sample ($n = 2{,}145$)	156.60(20)	.056 [.048, .065]	.953	.40 to .79
CFA with validation sample (with correlated errors; $n = 2{,}145$)	125.42(19)	.051 [.043, .060]	.963	.41 to .76

Note. df = degrees of freedom; RMSEA = root mean square error of approximation; CFI = comparative fit index; EFA = exploratory factor analysis; CFA = confirmatory factor analysis.
***All factor loadings are statistically significant at the .001 significance level.

Confirmatory Factor Analysis Results

We used a validation sample to validate the results of the one-factor solution suggested by EFA results. As shown in Table 4, results of the initial CFAs showed acceptable fit and were identical to the results of the EFAs. Conceptually, because we assumed that tests are nested in subjects, we expected to find a significant overlap between passing a test and passing different subjects. Thus, we allowed the errors between two items (i.e., Item 6, "How well do you succeed in passing all subjects?" and Item 8, "How well do you succeed in passing a test?") to correlate in subsequent CFA models for both baseline and follow-up. The CFA models with the correlated errors showed good fit at baseline ($\chi^2 = 93.93$, $df = 19$, $p < .001$, RMSEA = .043, 90% CI [.034, .052], CFI = .964) and follow-up ($\chi^2 = 125.42$, $df = 19$, $p < .001$, RMSEA = .051, 90% CI [.043, .060], CFI = .963). The factor loadings are listed in Table 3. The CFA factor loadings were generally consistent with EFA factor loadings; the widest gap between the CFA and EFA factor loadings was .05 at baseline and .04 at follow-up. The R^2 values of the eight items ranged from low to moderate at baseline (.16 to

Table 5
Results of Gender Invariance Tests at Baseline and Follow-Up

	χ^2(df)	RMSEA [90% CI]	CFI	ΔRMSEA	ΔCFI	Δχ^2(df)	p
Baseline							
CFA with male sample ($n = 1,060$)	64.64(19)***	.048 [.035, .061]	.960				
CFA with female sample ($n = 1,084$)	81.14(19)***	.055 [.043, .068]	.943				
Configural invariance	145.79(38)***	.051 [.043, .060]	.951				
Metric invariance	150.89(45)***	.047 [.039, .055]	.952	.004	.001	3.21(7)	.87
Scalar invariance	164.26(52)***	.045 [.037, .053]	.949	.002	.003	10.42(7)	.17
Follow Up							
CFA with male sample ($n = 1,060$)	64.51(19)***	.048 [.035, .061]	.971				
CFA with female sample ($n = 1,084$)	52.42(19)***	.040 [.028, .053]	.977				
Configural invariance	116.89(38)***	.044 [.035, .053]	.974				
Metric invariance	126.59(45)***	.041 [.033, .050]	.973	.003	.001	6.97(7)	.43
Scalar invariance	134.69(52)***	.039 [.031, .047]	.973	.002	.000	5.29(7)	.62

Note. df = degrees of freedom; RMSEA = root mean square error of approximation; CFI = comparative fit index;
EFA = exploratory factor analysis; CFA = confirmatory factor analysis; Δ = (delta) change in value.
*** $p < .001$.

.51) and at follow-up (.17 to .58). The Cronbach's alphas for the one-factor ASES were acceptable at baseline ($\alpha = .74$) and follow-up ($\alpha = .79$; George & Mallery, 2003). These results suggest not only that the scale has high internal consistency but that the scale's consistency is stable over time.

Gender Invariance at Baseline

To test for gender invariance at baseline, separate measurement models for boys and girls were estimated. Table 5 presents the results of all invariance tests. The boys-only model had a good fit with the data ($\chi^2 = 64.64$, $df = 19$, $p < .001$, RMSEA = .048, 90% CI [.035 – .061], CFI = .960) and the factor loadings (.39 to .62) were above the recommended ≥ .30 cutoff. Similarly, the girls-only model had an acceptable model fit with the data ($\chi^2 = 81.14$, $df = 19$, $p < .001$, RMSEA = .055, 90% CI [.043, .068], CFI = .943). The factor loadings of .38 to .69 were above the ≥ .30 cutoff, similar to those of the boys-only model (see Table 3). Following the adequacy of the one-factor model for the two gender groups, we proceeded with the invariance testing (Bowen & Masa, 2015).

First, all parameters were freed and the configural model yielded acceptable results ($\chi^2 = 145.79$, $df = 38$, $p < .001$, RMSEA = .051, 90% CI [.043, .060], CFI = .951). The results from the configurally invariant model confirmed that the boy and girl samples had the same factor structure. In the next step, the factor loadings were constrained to be equal between boys and girls. With constraints on the factor loadings, the metric model showed good fit ($\chi^2 = 150.89$, $df = 45$, $p < .001$, RMSEA = .047, 90% CI [.039 – .055], CFI = .952). A chi-square test of difference between the configural and metric models was not statistically significant ($\Delta\chi^2 = 3.21$, $\Delta df = 7$, $p = .87$), suggesting metric invariance between boys and girls. Stated differently, the factor loadings were equivalent between groups.

In the next restrictive model (the scalar model), all intercepts and factor loadings were held equal across boys and girls. The scalar model yielded a good fit with the data ($\chi^2 = 164.26$, $df = 52$, $p < .001$, RMSEA = .045, 90% CI [.037, .053], CFI = .949). A chi-square difference test between the metric and scalar models was not statistically significant ($\Delta\chi^2 = 10.42$, $\Delta df = 7$, $p = .17$), suggesting that the same self-efficacy construct was being measured across genders. In other words, for each academic self-efficacy score, boys and girls had similar intercepts on the eight scale items.

Gender Invariance at Follow-Up

To test for equivalence of the one-factor model between boys and girls at follow-up, we used the follow-up data to replicate the baseline invariance tests. First, we performed separate measurement models for boys and girls using only the validation sample at follow-up. The one-factor model of boys fit the data well: ($\chi^2 = 64.51$, $df = 19$, $p < .001$, RMSEA = .048, 90% CI [.035, .061], CFI = .971). All item loadings were statistically significant ($p < .001$), and they ranged from .36 to .70, which were above the $\geq$ .30 cutoff. The same one-factor model was replicated with girls at follow-up. The results showed a good fit with the data ($\chi^2 = 52.42$, $df = 19$, $p < .001$, RMSEA = .040, 90% CI [.028 – .053], CFI = .977). The factor loadings at follow-up (i.e., .33–.75) were statistically significant and above the $\geq$ .30 cutoff.

Following confirmation of adequate model fit with the separate boy and girl samples at follow-up, we ran three nested models (configural, metric, and scalar invariance). The configural model (with all parameters freed) showed excellent fit with the data ($\chi^2 = 116.89$, $df = 38$, $p < .001$, RMSEA = .044, 90% CI [.035, .053], CFI = .974). These results suggest that boys and girls had the same factor structure at follow-up. The metric model also had a good fit ($\chi^2 = 126.59$, $df = 4551$, $p < .001$, RMSEA = .041, 90% CI [.033, .050], CFI = .973). The chi-square difference test between the configural and metric models yielded statistically nonsignificant results ($\Delta\chi^2 = 6.97$, $\Delta df = 7$, $p = .43$). This finding suggests that the data support the hypothesis of equivalent factor loadings for boys and girls at follow-up. We proceeded with a test of scalar invariance by comparing the metric model to a scalar model ($\chi^2 = 134.69$, $df = 52$, $p < .001$, RMSEA = .039, 90% CI [.031, .047], CFI = .973). The chi-square difference test between the metric and scalar models showed that the two models were not significantly different at follow-up ($\Delta\chi^2 = 5.29$, $\Delta df = 7$, $p = .62$). This result suggests that the one-factor model had strong gender invariance at follow-up. This finding means that boys and girls had similar intercepts on each of the eight items of academic self-efficacy at follow-up.

Discussion

The goal of this study was to investigate (a) the construct validity of the ASES in the Ghanaian context, (b) the measurement equivalence of the scale across gender, and (c) the temporal stability of the gender equivalence. The results support Muris' (2001) finding of a one-dimensional ASES. Findings of this study also provide empirical support for the generalizability of the scale to boys and girls at two points in time.

Because academic self-efficacy includes a range of activities connected to academic work, one of the goals of this study was to examine the factor structure of the academic self-efficacy construct. A previous U.S.-based study by Muris (2001) tested the construct as unidimensional. To confirm Muris' hypothesis of a one-dimensional construct, we ran separate CFAs using data collected at two time points with a 3-year interval. Results from both time points suggest that the one-dimensional construct is supported by the data from Ghana. The fact that this one-dimensional finding was consistent across multiple random subsamples (i.e., calibration and validation samples, and baseline and follow-up samples, as well as boy and girl samples) lends firm support for the one-dimensional structure of the academic self-efficacy construct. This finding suggests that education and psychosocial researchers in developing countries like Ghana could potentially conceptualize academic self-efficacy in line with Muris' theory of a one-dimensional construct that measures young people's perceived ability to manage their learning behavior, master academic subjects, and fulfill academic expectations.

Despite the consistency in the factor structure between our findings and those of Muris, our findings differ from those of Muris in three important ways. First, our factor loadings were generally

low with wide variability. Second, our study retained all eight scale items in the final model because all items met the factor loadings cutoff criteria (i.e., > .30) and were statistically significant. In contrast, Muris' study dropped one item ("How well can you get teachers to help you when you get stuck on schoolwork?") from the final factor analysis because it did not load substantially. We think retaining this item in the analysis is important, given the item's relevance in collective-oriented cultures in Africa, Asia, Central and South America, and the Middle East (Woodward & Denton, 2013), where students are comparatively less assertive and often find it difficult to approach teachers because of the power differential (Joy & Kolb, 2009). In Ghana and many sub-Saharan African contexts, students are expected to show unqualified respect for teachers (Whitehead, 2007). This social expectation has implications for the student-teacher relationship because it might affect the student's confidence in seeking assistance from the teacher when the student struggles with academic work. Students with the ability to overcome such stressful relationships might find it easier to ask for and receive help from teachers with challenging academic tasks. The third difference between our findings and those reported by Muris is that in the present study, we allowed the errors between two items ("How well do you succeed in passing all subjects?" and "How well do you succeed in passing a test?") to be correlated. We modeled the construct this way because we believe that students who have encountered difficulties in passing subject tests might struggle to succeed in their courses more generally.

This study's findings have implications for further testing of the ASES and the adaptation of other scales to culturally diverse settings, including the increasing diverse classrooms in Western contexts. The items and expanded response set met three key criteria (i.e., magnitude, strength, and generality) essential for judging the adequacy of self-efficacy scales (Lunenburg, 2011; Van der Bijl & Shortridge-Baggett, 2002). Consistent with Van der Bijl and Shortridge-Baggett's (2002) recommendations, this study used an expanded response set to improve response variability. The wide variability of the response set allowed respondents to indicate the magnitude of the efforts that they felt was required to perform each of the eight specific tasks. Moreover, from an evaluative perspective, expansion of the original 5-point response scale to an 11-point scale was an improvement worth exploring in other measurement scales. Comparatively, the wider scale is more sensitive and able to detect subtle differences compared with the original scale with fewer response options. Without compromising the integrity of the academic self-efficacy construct as unidimensional, the expanded response set offered greater statistical reliability and validity.

Although some of the items in the ASES might seem redundant on casual assessment, the set of items is essential to capture the extent to which young people can perform tasks of different levels of difficulty, as recommended in the measurement of self-efficacy (Van der Bijl & Shortridge-Baggett, 2002). For instance, one item measured how well respondents could "pass a test" and another item asked students how well they could "pay attention in class." Arguably, passing a test is a task at a higher level of difficulty than paying attention in class because passing a test usually requires preparation and is likely to be a more stressful experience for the student.

In terms of meeting the criteria of scale generality (Lunenburg, 2011), the adapted ASES assesses expectations across multiple situations related to academic work. Specifically, the ASES items cover at least five distinct subdomains: (a) getting help, (b) studying, (c) finishing homework, (d) passing a test/subject, and (e) paying attention in class—all of which are related to academic work. The array of academic-related activities covered by the scale is further indication of the scale's relevance as a valid measure of self-efficacy in the academic domain.

Results of tests of gender invariance at two measurement occasions have important implications for the use of the scale with boys and girls. A reliable ASES must be sensitive enough to capture changes in self-efficacy regardless of personal characteristics. A scale that reflects false changes based on scale sensitivity to personal characteristics such as gender, rather than reflecting true

changes based on actual differences, has limited, if any, utility for mixed gender samples because the scale would measure different issues for different groups. Results of the present study suggest the adapted scale has wide application and may be used with both boys and girls. In addition, the gender neutrality of the latent construct is supported by the present study's finding of temporal stability of measurement invariance across gender. These findings mean that researchers do not have to invest time or resources in developing gender-specific instruments, thereby addressing some of the financial challenges to the widespread use of the validated instruments within the Ghanaian community and other resource-limited countries.

Furthermore, based on this study's findings, the construct of academic self-efficacy is gender invariant even after 3 years. This finding means the ASES should be effective for assessing true changes in young people's academic self-efficacy without worry about evolving gender roles at the adolescent stage. Without an instrument's demonstrated gender neutrality, adolescents' evolving gender roles might compromise the ability to measure the same self-efficacy attributes among boys and girls over time. Thus, the temporal stability finding further legitimizes the use of the adapted eight-item ASES in experimental and intervention studies that seek to assess the impacts and predictors of academic self-efficacy.

Moreover, the fact that the items underlying the academic self-efficacy construct did not require extensive changes in adapting from a Western to a non-Western context has implications for the use of the ASES in multicultural settings, including schools in the United States, which are projected to become even more ethinically and culturally diverse. The successful validation of the ASES also has implications for the testing of structural relationships in future studies. If future measurement studies of academic self-efficacy yield consistent results, such findings will serve as an important prelude to testing structural relationships in culturally and ethnically diverse settings, including the effects of educational interventions on academic self-efficacy, or the effects of academic self-efficacy on school-related outcomes, such as academic achievement and in-school behavior. Well-established measurement models for academic self-efficacy can reduce the likelihood of finding poor-fitting structural solutions when testing structural relationships.

Further, the protocols used in adapting the ASES to the Ghanaian context offer important insights into ways in which other psychosocial scales can be normed and tested for use in different sociocultural contexts, including increasingly diverse classrooms in the United States. For example, strategies used in this study to adapt a measurement scale for a different sociocultural context included the use of an expert panel, the expansion of the response set, and then the subsequent pilot test that used multiple methods for data collection; these strategies could be explored in future studies with similar goals. In addition, the findings of this study have implications for the use of the validated scale in educational training and development practice. Developing site-specific instruments that provide reliable and valid measures of academic self-efficacy is an undertaking that requires technical skills, resources, and time beyond that available to most education professionals, such as educational development specialists, practitioners, and administrators. Thus, these and other education stakeholders will be able to use the validated ASES to obtain reliable and valid information about the academic self-efficacy level of Ghanaian youth and youth from other developing countries that share similar characteristics with Ghana.

CONCLUSION

The lack of measurement scales for assessing academic self-efficacy and other psychosocial outcomes among Ghanaian youth has resulted in a huge gap in the empirical literature. The academic self-efficacy construct is conspicuously missing in education and youth development studies in Ghana because of, at least in part, the lack of validated measurement scales. For this reason, adapting and

validating an ASES to the Ghanaian context enhances the understanding of education and youth development research and practice.

The results of this study must be appraised in light of the study limitations. First, the study did not assess longitudinal invariance of the adapted ASES. In addition, the study sample did not include youth from the Upper East and Upper West regions of Ghana, which limits our ability to generalize results to those regions. Nonetheless, the CFA results suggest the adapted ASES has utility for educational and psychosocial research in developing countries like Ghana. The scale is relatively short in length and parsimonious. Thus, this adapted scale offers researchers and practitioners who work in culturally diverse environments a viable option for assessing students' academic self-efficacy. Moreover, the finding of gender invariance suggests that the ASES measures the same academic self-efficacy traits in boys and girls, and therefore, the scale can be used confidently with mixed-gender samples. However, further psychometric research is needed to test invariance of the scale across groupings (i.e., beyond culture and gender) to enhance the potential use of the scale in heterogeneous populations, such as studies comparing academic self-efficacy of rural versus urban youth.

References

Alwin, D. F. (1997). Feeling thermometers vs 7-point scales. Sociological Methods and Research, 25, 318–351.

Asante, K. O., & Doku, P. N. (2010). Cultural adaptation of the Condom Use Self Efficacy Scale (CUSES) in Ghana. BMC Public Health, 10, 1–7.

Atindanbila, S., Winifred, A. D., & Awuah-Peasah, D. (2012). The effects of parenting on the self-esteem of adolescents: A study at Labadi Presbyterian Secondary School (Ghana). Research on Humanities and Social Sciences, 2, 13–18.

Baah-Odoom, D., & Riley, G. A. (2013). The role HIV-related blame and stigmatisation play on risk perception, self-efficacy and sexual behaviour among students in Ghana. IFE Psychologia, 21, 284–303.

Babakus, E., Ferguson, C. E., Jr., & Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. Journal of Marketing Research, 24, 222–228.

Bandura, A. (1986). Social foundations of thought and action. Englewood Cliffs: NJ. Prentice Hall.

Bandura, A., Pastorelli, C., Barbaranelli, C., & Caprara, G. V. (1999). Self-efficacy pathways to childhood depression. Journal of Personality and Social Psychology, 76, 258–269.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. Spine, 25, 3186–3191.

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. Public Opinion Quarterly, 71, 287–311.

Borsa, J. C., Damásio, B. F., & Bandeira, D. R. (2012). Cross-cultural adaptation and validation of psychological instruments: Some considerations. Paidéia (Ribeirão Preto), 22, 423–432.

Bowen, N. K., & Masa, R. (2015). Conducting measurement invariance tests with ordinal data: A guide for social work researchers. Journal of the Society for Social Work and Research, 6, 229–249.

Britner, S. L., & Pajares, F. (2001). Self-efficacy beliefs, motivation, race, and gender in middle school science. Journal of Women and Minorities in Science and Engineering, 7, 269–279.

Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. Journal of Research in Science Teaching, 43, 485–499.

Bryant, F. B., & Satorrab, A. (2012). Principles and practice of scaled difference chi-square testing. Structural Equation Modeling: A Multidisciplinary Journal, 19, 372–398.

Bussey, K. (2011). The influence of gender on students' self-regulated learning and performance. In B. J. Zimmerman & D. H. Schunk (Eds.), Handbook of self-regulation of learning and performance (pp. 408–441). New York, NY: Routledge.

Chemers, M. M., Hu, L. T., & Garcia, B. F. (2001). Academic self-efficacy and first year college student performance and adjustment. Journal of Educational Psychology, 93, 55–64.

Chowa, G. A. N., Masa, R. D., Ansong, D., & Ramos, Y. (2015). How do student and school characteristics influence youth academic achievement in Ghana? A hierarchical linear modeling of Ghana YouthSave baseline data. International Journal of Educational Development, 45, 129–140.

Cizek, G. J. (2012). Defining and distinguishing validity: Interpretations of score meaning and justifications of test use. Psychological Methods, 17, 31.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical Assessment Research & Evaluation, 10, 1–9.

Dawes, J. G. (2002). Five point vs eleven point scales: Does it make a difference to data characteristics? Australasian Journal of Market Research, 10, 39–47.

Elliott, W., III, Sherraden, M. S., Johnson, L., Johnson, S., & Peterson, S. (2007). College expectations among young children: The potential role of savings (Working Paper No. 07-06). St. Louis, MO: Center for Social Development, Washington University.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. Psychological Bulletin, 136, 103–127.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods, 3, 272–299.

George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.). Boston, MA: Allyn & Bacon.

Gota, A. A. (2012). Effects of parenting styles, academic self-efficacy, and achievement motivation on the academic achievement of university students in Ethiopia (Unpublished doctoral dissertation, Edith Cowan University, School of Psychology and Social Science, Joondalup, Western Australia). Retrieved from http://ro.ecu.edu.au/theses/461

Huang, C. (2013). Gender differences in academic self-efficacy: a meta-analysis. European Journal of Psychology of Education, 28, 1–35. doi:10.1007/s10212-011-0097-y.

Hau, K.-T., & Marsh, H. W. (2004). The use of item parcels in structural equation modeling: Non-normal data and small sample sizes. British Journal of Mathematical and Statistical Psychology, 57, 327–351.

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

Imhonde, H. O. (2013). Self-esteem, gender, family-communication-style and parental neglect as predictors of aggressive tendencies among secondary school adolescents. International Journal of Public Health Science, 2, 93–100.

Joe, H., & Maydeu-Olivares, A. (2010). A general family of limited information goodness-of-fit statistics for multinomial data. Psychometrika, 75, 393–419.

Joy, S., & Kolb, D. A. (2009). Are there cultural differences in learning style? International Journal of Intercultural Relations, 33, 69–85.

Kline, R. B. (2005). Principles and practice of structural equation modeling. New York, NY: Guilford Press.

Kling, K. C. (1999). Gender differences in self-esteem: A meta-analysis. Psychological Bulletin, 125, 470–500.

Lunenburg, F. (2011). Self-efficacy in the workplace: implications for motivation and performance. International Journal of Management, Business, and Administration, 14, 1–7. Retrieved from http://www.nationalforum.com/Electronic%20Journal%20Volumes/Lunenburg,%20Fred%20C.%20Self-Efficacy%20in%20the%20Workplace%20IJMBA%20V14%20N1%202011.pdf.

Matoti, S. H. (2011). Measuring the academic self-efficacy of students at a South African higher education institution. Journal of Psychology in Africa, 21, 151–154.

Maydeu-Olivares, A., & Joe, H. (2005). Limited- and full-information estimation and goodness-of-fit testing in 2n contingency tables: A unified framework. Journal of the American Statistical Association, 100, 1009–1020.

Millsap, R. E., & Olivera-Aguilar, M. (2012). Investigating measurement invariance using confirmatory factor analysis. In R. H. Hoyle (Ed.), Handbook of structural equation modeling (pp. 380–392). New York, NY: Guilford Press.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. Journal of Counseling Psychology, 38, 30–38.

Muris, P. (2001). A brief questionnaire for measuring self-efficacy in youths. Journal of Psychopathology and Behavioral Assessment, 2(3), 145–149.

Muthén, L. K., & Muthén, B. O. (1998–2012). Mplus user's guide (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nunn, J. S., & Thomas, S. L. (1999). The angry male and the passive female: The role of gender and self-esteem in anger expression. Social Behavior & Personality, 27, 145–154.

Nyarko, K. (2011). Parental school involvement: The case of Ghana. Journal of Emerging Trends in Educational Research and Policy Studies, 2, 378–381.

Oettingen, G. (1997). Cross-cultural perspectives on self-efficacy. In A. Bandura (Ed.), Self-efficacy in changing societies (pp.149–157). New York, NY: Cambridge University Press.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. Psychometrika, 44, 443–460.

Scholz, U., Doña, B. G., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. European Journal of Psychological Assessment, 18, 242–251.

Schunk, D. H. (1985). Self-efficacy and classroom learning. Psychology in the Schools, 22, 208–223.

Schunk, D., & Usher, E. (2011). Assessing self-efficacy for self-regulated learning. In B. Zimmerman & D. H. Schunk (Eds.), Handbook of self-regulation of learning and performance (pp. 282–297). New York, NY: Taylor & Francis.

Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. Review of Educational Research, 78, 751–796.

Van der Bijl, J. J., & Shortridge-Baggett, L. M. (2002). The theory and measurement of the self-efficacy construct. In E. A. Lentz & L. M. Shortridge-Baggett (Eds.), Self-efficacy in nursing: Research and measurement perspectives (pp. 9–28). New York, NY: Springer.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organizational Research Methods, 3, 4–69.

Whitehead, D. M. (2007). Taking the road less traveled: Primary teacher retention in Ghana (Doctoral dissertation, Indiana University). Dissertation Abstracts International, 68-09, 3797A.

Woodward, G. C., & Denton, R.E. (2013). Persuasion and influence in American life (7th ed.). Long Groove, IL: Waveland Press.

Wu, Q., Tsang, B., & Ming, H. (2014). Social capital, family support, resilience and educational outcomes of Chinese migrant children. British Journal of Social Work, 44, 636–656.

Yan, W., & Gaier, E.L. (1994). Causal attributions for college success and failure: An Asian-American comparison. Journal of Cross-Cultural Psychology, 25, 146–158.

Yendork, J. F., & Somhlaba, N. Z. (2015). Do social support, self-efficacy and resilience influence the experience of stress in Ghanaian orphans? An exploratory study. Child Care in Practice, 21, 140–159.