



Piloting a Short Form of the Academic Competence Evaluation Scales

Christopher J. Anthony¹ · James C. DiPerna²

Published online: 13 February 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

A growing body of research indicates that noncognitive factors are important predictors of students' academic and life success (e.g., Garcia, The need to address noncognitive skills in the education policy agenda (Briefing Paper No. 386), <http://files.eric.ed.gov/fulltext/ED558126.pdf>, 2014). Despite this evidence base, there are few psychometrically sound measures of such factors appropriate for use in research and practice. One currently available measure is the Academic Competence Evaluation Scales (ACES; DiPerna and Elliott, Academic Competence Evaluation Scales, The Psychological Corporation, San Antonio, TX, 2000) which assesses the skills, attitudes, and behaviors of students that contribute to school success. The length of the ACES (73 items) may limit its use at the primary and secondary levels within a multi-tiered service delivery system or for large-scale educational research. To address this need, the current study piloted a short form of the ACES (ASF) with a sample of 301 elementary students. Results provided initial evidence for the reliability and validity of scores from the ASF.

Keywords Noncognitive factors · Assessment · Academic enablers · Academic skills · Academic Competence Evaluation Scales

Introduction

During the past decade, a body of research has emerged regarding students' "noncognitive"¹ factors related to academic and life success (Farrington et al., 2012; Garcia, 2014). Much of this research has indicated that these variables are related to academic achievement (Farrington et al., 2012) as well as a number of other important life outcomes such as earnings (Garcia, 2014). As a result, many education stakeholders have identified these factors as important educational outcomes (e.g., Zeehandelaar & Winkler, 2013), and a number of researchers (Farrington et al., 2012; Garcia, 2014; Rosen, Glennie, Dalton, Lennon, & Bozick, 2010) have called for a greater focus on noncognitive factors in educational research, policy, and practice.

The increasing interest regarding noncognitive factors also has exposed several important limitations to the rigor

of scientific inquiry in this domain (Duckworth & Yeager, 2015). One challenge limiting progress in research and practice has been that different stakeholders have used different terms and frameworks to characterize the skills, attitudes, and behaviors most often associated with the noncognitive domain. In response to this challenge, the University of Chicago Consortium on Chicago School Research (CCSR; Farrington et al., 2012) developed a framework specifying five domains of noncognitive factors: academic behaviors, academic perseverance, academic mindsets, learning strategies, and social skills. These domains are promising because evidence suggests that they are directly and indirectly related to academic achievement as well as other life outcomes (e.g., Borghans, Duckworth, Heckman, & Ter Wheel, 2008; Heckman, Stixrud, & Urzua, 2006) and are malleable (e.g., Kautz, Heckman, Diris, Ter Wheel, & Borghans, 2014). Thus, they represent prime targets for intervention and prevention programs.

Another significant limitation is the lack of psychometrically sound measures available to assess noncognitive factors

✉ Christopher J. Anthony
christopher.anthony@okstate.edu

¹ School of Teaching, Learning, and Educational Sciences, Oklahoma State University, 223 Willard Hall, Stillwater, OK 74078, USA

² The Pennsylvania State University, University Park, PA, USA

¹ We share the various concerns raised elsewhere (e.g., Farrington et al., 2012) regarding the appropriateness of using the term "noncognitive" when referring to these variables; however we decided to retain noncognitive here because it is the most commonly used term for these variables.

(e.g., Credé, Tynan, & Harms, 2017; Duckworth & Yeager, 2015; West et al., 2016). Though developed and published before the emergence of the “noncognitive” label, the Academic Competence Evaluation Scales (ACES; DiPerna & Elliott, 2000) is one measure that assesses a number of the constructs identified in the CCSR framework (Farrington et al., 2012). The four academic enablers subscales of the ACES (interpersonal skills, engagement, motivation, and study skills) are consistent with four of the five noncognitive factor domains (social skills, academic behaviors, academic perseverance, and learning strategies) in the CCSR framework. The construct definitions across the ACES and CCSR frameworks also are similar. For example, DiPerna and Elliott (2000) defined study skills as “behaviors or strategies that facilitate the processing of new material” (p. 7). Similarly, Farrington et al. (2012) defined learning strategies as “processes and tactics one employs to aid in the cognitive work of thinking, remembering, or learning” (p. 10). Although the ACES does not assess the CCSR academic mindsets domain, the significant overlap between the two frameworks provides independent support for the ACES as a measure of several “noncognitive” domains.

In addition to its overlap with the CCSR framework, the ACES produces scores with psychometric evidence to support their use (e.g., DiPerna & Elliott, 2000; Hambleton, 2010; Sabers & Bonner, 2010). With regard to reliability evidence, internal consistency estimates from the ACES have been high (i.e., $> .90$ except for the test–retest reliability coefficient for the Critical Thinking subscale, which was $.88$) across all scales and subscales of the ACES (DiPerna & Elliott, 2000). Additionally, scores from the ACES have been shown to relate as expected with scores from measures of related constructs such as the Social Skills Rating System (Gresham & Elliott, 1990) and the Wechsler Individual Achievement Test—Second Edition (Wechsler, 2002). Finally, results of exploratory factor analyses have provided support for the structural validity of the ACES, specifically a correlated factors model (DiPerna & Elliott, 2000).

Given its evidence base and the constructs assessed, the ACES has been used to inform intervention planning and outcome evaluation in research (e.g., Volpe et al., 2006; Demaray & Jenkins, 2011; McCormick, O’Connor, Cappella, & McClowry, 2013) and practice (Cleary, Gubi, & Prescott, 2010). The published teacher form of the measure (ACES-Teacher Form; ACES-TF), however, includes 73 items and requires approximately 15–20 min to complete, which may pose a challenge for using the measure at the primary and secondary levels within a multi-tiered service delivery system (Brady, Evans, Berlin, Bunford, & Kern, 2012) or for large-scale educational research. Such limitations could be addressed; however, by the development of a

short form of the ACES that is more efficient yet maintains the original structure of the measure.

To address this need, Anthony & DiPerna (2017) identified a set of maximally efficient items (SMI) for each ACES-TF subscale using item response theory (IRT) and procedures recommended by Smith, McCarthy, and Anderson (2000). Despite initial evidence for the psychometric adequacy of SMI scores (Anthony & DiPerna, 2017), data were from a single administration of the full-length ACES. Although information gleaned using such an approach can be an important initial step for short form development, this methodology is insufficient to substantiate use of short forms (Smith et al., 2000).

Although the creation of short forms is common, the resulting measures often are limited due to a number of problematic practices (Credé, Harms, Niehorster, & Gaye-Valentine, 2012; Smith et al., 2000). For example, researchers frequently derive short forms through modifying existing measures, but they do not commonly report psychometric properties of the shortened measures (Smith et al., 2000). In the domain of social competence, for instance, Zaslou et al. (2006) found that 27% of studies published from 1979 to 2005 modified extant measures without reporting psychometric evidence for the abbreviated measures. Additional problems in short form development include using insufficiently validated parent measures to create short forms, failing to use independent administrations for short form validation studies, and failing to show that short forms retain the factor structures of their parent measures (Smith et al., 2000).

As outlined by Smith et al. (2000), there are several key steps to validating short form measures. First, it is important to independently administer short forms for validation studies (rather than merely examine properties of sets of items drawn from a single administration of a parent form). When independently administered short form data are acquired, Smith et al. noted several important pieces of information necessary to substantiate use and interpretation of short form scores. First, these authors emphasized the examination of subscale reliability coefficients to ensure that the short form development process has not led to unacceptable degradation of score reliability. Next, Smith et al. noted that it is important to provide evidence that short forms retain the factor structure of their parent measures. Finally, concurrent validity evidence is crucial for establishing the construct validity of scores from any measure (APA, AERA, NCME, 2014) and is especially important for validating short forms, as it cannot be assumed that short forms retain the psychometric properties of their parent forms (Smith et al., 2000). Given that all SMI evidence was gleaned from a single administration of the full-length ACES-TF in the Anthony & DiPerna (2017) study, the primary purpose of this study

was to examine the initial psychometric properties of a short form of the ACES-TF (the ACES—Short Form; ASF).

Related to these goals, we tested several hypotheses. First, we hypothesized that the structure of the ASF would be consistent with the structure of the ACES-TF (DiPerna & Elliott, 2000). Second, we predicted that scores from the ASF would be associated with reliability coefficients acceptable for individual decision-making. Third, we tested a series of convergent validity hypotheses (APA, AERA, NCME, 2014). Based on previous findings with the full-length ACES (e.g., DiPerna & Elliott, 2000) we predicted that ASF Academic Skills scales would produce moderate to large relationships with directly measured academic achievement. Also, informed by research examining the relationship between social skills and academic skills (e.g., Malecki & Elliott, 2002) we predicted that ASF Academic Skills scales would demonstrate moderate positive relationships with teacher-rated social skills and moderate negative relationships with teacher-rated problem behaviors. Based on prior evidence (DiPerna & Elliott, 2000), we also predicted that ASF Academic Enabler scales would be moderately associated with directly measured academic achievement. Finally, we predicted that ASF Academic Enabler scales would produce large positive relationships with teacher-rated social skills and large negative relationships with teacher-rated problem behaviors.

Method

Participants

Students and teachers from 7 schools and 63 elementary classrooms were invited to participate in the project. Teachers initially received a written description of the study along with a consent form. After a teacher agreed to participate, an invitation letter and consent form were sent to the parents of each child in the teacher's classroom. A reminder letter then was distributed to parents approximately 1 week after receipt of the initial communication. Prior to their participation, students with parental consent were provided with a brief verbal explanation of the project and asked if they wanted to participate. Students who provided assent were then included in the study.

As shown in Table 1, the sample consisted of 301 second through sixth-grade students² with a median age of 8.83 years (range 6.67–12.33 years). With regard to grade, 22% of students were in second grade, 26% in third grade, 23% in fourth grade, 16% in fifth grade, and 13% in sixth

Table 1 Demographic characteristics (percentages) of participants ($N = 301$) and corresponding national estimates

Characteristic	Participants	US student population ^a
Gender		
Female	50	51
Male	49	49
Race/ethnicity		
White	77	50
Black	12	16
Hispanic	6	25
Other	2	9
Educational status		
General education	90	86
Special education	9	14

Percentages do not sum to 100 in some cases due to missing data and/or rounding

^aAll national figures drawn from the 2013 to 2014 Civil Rights Data Collection survey of public schools (U.S. Department of Education Office for Civil Rights, 2016)

grade. Teachers were predominately female (85%), white (98%), had a bachelor's degree (79%), and had extensive teaching experience (median = 15.5 years).

Measures

Academic Competence Evaluation Scales-Short Form (ASF)

The focal measure for this study consisted of an independently administered short form of the ACES-TF (the ASF) including a set of 32 maximally efficient items (SMIs) identified by Anthony & DiPerna (2017). Consistent with its parent version, the ASF includes three Academic Skills scales (Reading, Mathematics, and Critical Thinking) and four Academic Enablers scales (Interpersonal Skills, Engagement, Motivation, and Study Skills). All items are rated on a 5-point Likert scale ranging from 1 (*Never*) to 5 (*Almost Always*). Anthony & DiPerna (2017) examined Test Information Functions (TIFs) to evaluate reliability for each scale SMI. Across broad ranges of theta (the latent trait being measured) SMI scores produced information values greater than a .90 reliability standard. Despite initial evidence regarding score reliability, validity of scores from these SMIs has not been examined previously and is the primary focus of this study.

STAR Reading and Math

The STAR Reading (Renaissance Learning, 2015) and Math (Renaissance Learning, 2012) assessments are computer adaptive tests designed to assess the reading and math skills

² After the deletion of one outlier identified through calculation of Mahalanobis distances with the total sample.

of students across first through twelfth grades. The STAR Reading test focuses on skill such as word knowledge, comprehension strategies, and analysis of text. The STAR Math test measures student skills in such topic domains as numbers and operations, measurement, and geometry. Overall reliability coefficients for STAR Reading scores ranged from .89 to .91 for second through sixth-grade students from the standardization sample. For STAR Math scores, reliability coefficients were somewhat lower (.79–.84 across second through sixth-grade students from the standardization sample, though still adequate for research purposes (Salvia, Ysseldyke, & Bolt, 2010). Based on a synthesis of concurrent and predictive validity coefficients from STAR validity studies with similar academic measures (Renaissance Learning, 2012, 2015), overall validity coefficients range from .77 to .78 for STAR Reading scores and from .63 to .72 for STAR Math scores for students in the second through sixth grade.

Social Skills Improvement System-Teacher Rating Scales

The Social Skills and Problem Behaviors scales and subscales of the Social Skills Improvement System-Teacher Rating Scale (SSIS-TRS; Gresham & Elliott, 2008) also were collected in this study. As reported in the technical manual, there is evidence for the reliability and validity of scores from the SSIS-TRS (Gresham & Elliott, 2008). With regard to reliability, Cronbach's α ranged from .78 to .97 (median = .97) and stability coefficients ranged from .68 to .86 (median = .82) across all scales and subscales for the standardization sample. As evidence for validity, scores from the SSIS-TRS correlated as expected with scores from various measures (e.g., the Behavioral Assessment System for Children—Second Edition; Reynolds & Kamphaus, 2004) both in the standardization sample (Gresham & Elliott, 2008) and in subsequent independent research (e.g., Gresham, Elliott, Cook, Vance, & Kettler, 2010; Gresham, Elliott, Vance, & Cook, 2011).

Procedures

Data were collected at the conclusion of a multi-year project evaluating the efficacy of the Social Skills Improvement System-Classwide Intervention Program (SSIS-CIP; Elliott & Gresham, 2007). Seven schools participated in this study and classrooms were randomly assigned to treatment and control groups. In the final year of the larger study, SSIS-TRS data were collected for all students. Due to resource constraints, the STAR Reading and Mathematics tests were administered to a random subsample of students stratified by gender. As a result, though teachers participating in this study provided ASF ratings for all participating students in their classrooms,

only a subsample of participating students had achievement data ($n = 162$ for reading, $n = 159$ for math).³ Social skills, problem behaviors, and academic data were collected during the latter part of the school year (late February–early April). Teachers then completed the ASF during a separate data collection window in the last month (May–June) of the school year. The average interval between ASF and validity measures was approximately 10 weeks for the SSIS-TRS and 11 weeks for the STAR measures.

Data Analysis

Several data analytic techniques were used to examine ASF scores. First, to evaluate structural validity, a confirmatory factor analysis (CFA) was conducted. Prior to conducting the CFA, data were screened to ensure they met underlying assumptions (e.g., outliers, normality). One outlier was identified through examination of Mahalanobis distances and leverage values (Field, 2009), and this case was subsequently deleted for all analyses. No significant skew or kurtosis values were observed for any ASF item. Thus, although item level data were ordinal, the robust maximum likelihood (MLR) estimator in MPlus 7 (Muthén & Muthén, 2012) was used for the CFA. Rhemtulla, Brosseau-Liard, and Savalei (2012) recommended this approach when there are more than four response options and smaller sample sizes.

The structure tested in this analysis was a correlated factors design in which each ASF scale (e.g., Reading/Language Arts and Engagement) was represented by a factor, and all factors were allowed to intercorrelate. This approach was selected because prior structural analyses of the ACES-TF were exploratory (e.g., DiPerna & Elliott, 2000) with oblique rotations consistent with a correlated factors design (Fig. 1). Model fit was evaluated relative to Hu and Bentler's (1999) recommended thresholds for the Root Mean Squared Error of Approximation (RMSEA; $\leq .06$), Comparative Fit Index (CFI; $\geq .95$), Tucker Lewis Index (TLI; $\geq .95$), χ^2 ($p > .05$), and the Standardized Root Mean Square Residual value (SRMR; $\leq .08$). Next, Cronbach's α values were calculated and examined for each ASF scale and the two ASF total scales (Academic Skills and Academic Enablers). Finally, convergent validity analyses consisted of computing correlations between ASF scores and SSIS-TRS and STAR scores. Based on Cohen's (1988) guidelines, correlations ($|r|$) were interpreted as small (.10–.30), moderate (.30–.50), or large ($> .50$).

³ After the deletion of three outliers detected by Mahalanobis distances calculated with this subsample.

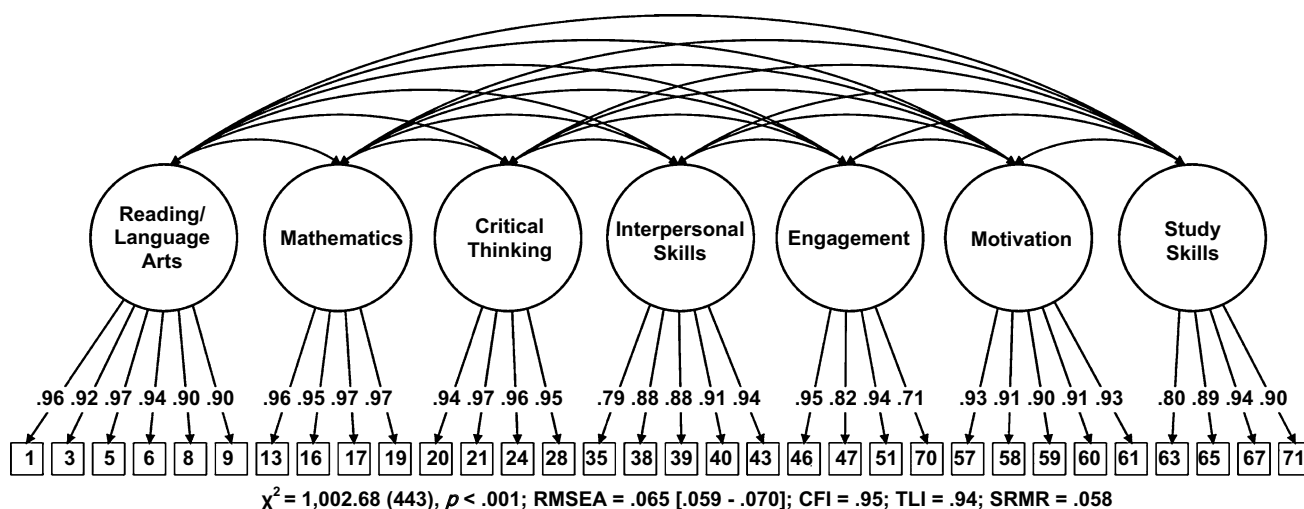


Fig. 1 Correlated factors confirmatory factor analysis model

Table 2 Interfactor correlations from the confirmatory factor analysis ($N = 301$)

	1	2	3	4	5	6	7
1. Reading/Language arts	–						
2. Mathematics	.86	–					
3. Critical thinking	.90	.89	–				
4. Interpersonal skills	.27	.24	.29	–			
5. Engagement	.49	.46	.53	.57	–		
6. Motivation	.57	.54	.59	.72	.78	–	
7. Study skills	.53	.46	.52	.69	.60	.79	–

Results

Initially, the CFA was conducted adjusting for the nested structure of the data (students nested within teacher). This approach generated MPLUS warnings because the number of parameters estimated (117) exceeded the number of available clusters (63). As such, the model was examined without the clustering adjustment, and the results from each method were compared. As there were no substantive differences between the models (e.g., loadings were identical, RMSEA, CFI, and TLI differed by .002 and SRMR values were identical across models), reported results are from the noncluster-adjusted model.

The χ^2 value of this model (Fig. 1) was statistically significant; $\chi^2 = 1002.68 (443), p < .001$. The RMSEA associated with this model was .065 (90% CI .059–.070) and the CFI and TLI values were .95 and .94, respectively. Finally, the SRMR value was .058. Standardized loadings of items (Fig. 1) on their corresponding factors were high, ranging from .90 to .97 (median = .96) for Academic Skills items and from .71 to .95 (median = .91) for Academic Enablers items. Interafactor correlations (Table 2)

between Academic Skills factors ranged from .86 to .90 (median = .89). Interafactor correlations between Academic Enablers factors ranged from .57 to .79 (median = .71). Finally, interfactor correlations between Academic Skills and Academic Enablers factors ranged from .24 to .59 (median = .51).

To examine ASF score reliability, Cronbach's α was computed for all ASF scales (Table 3). Estimated reliability was high for all scales. Specifically, Academic Skills scales all produced α values of .98 and Academic Enablers scales produced α values ranging from .91 to .96. Correlations were also computed to evaluate convergent validity (Table 3). ASF Academic Skills scale scores generally demonstrated large positive relationships with STAR Reading and Mathematics scores ($.47 \leq r \leq .56$), moderate positive relationships with SSIS-TRS Social Skills scores ($.24 \leq r \leq .33$) and moderate negative relationships with SSIS-TRS Problem Behaviors scores ($-.31 \leq r \leq -.27$). ASF Academic Enablers scale scores generally yielded small to moderate positive relationships with STAR Reading and Mathematics scores ($.18 \leq r \leq .40$), large positive relationships with SSIS-TRS Social Skills scores

Table 3 Cronbach's alpha and correlations between ASF and validity measures

ASF Scale	Cronbach's α	Social Skills Improvement System-Teacher Rating Scale		STAR	
		Social skills	Problem behaviors	Reading	Mathematics
Academic skills					
Reading/Language arts	.98	.29 [.18, .39]	-.28 [-.38, -.17]	.56 [.44, .66]	.51 [.38, .62]
Mathematics	.98	.24 [.13, .34]	-.27 [-.37, -.16]	.51 [.39, .62]	.47 [.34, .58]
Critical thinking	.98	.33 [.23, .43]	-.31 [-.41, -.20]	.55 [.43, .65]	.52 [.40, .63]
Academic enablers					
Interpersonal skills	.94	.78 [.73, .82]	-.73 [-.78, -.67]	.21 [.06, .35]	.29 [.14, .43]
Engagement	.91	.60 [.52, .67]	-.48 [-.56, -.39]	.18 [.03, .33]	.26 [.11, .40]
Motivation	.96	.67 [.60, .73]	-.59 [-.66, -.51]	.37 [.23, .50]	.40 [.26, .52]
Study skills	.93	.60 [.52, .67]	-.54 [-.62, -.45]	.28 [.13, .42]	.35 [.21, .48]

$N = 301$ for Cronbach's α and Social Skills Improvement System-Teacher Rating Scale correlations; $N = 162$ and 159 for STAR Reading and Mathematics correlations, respectively

($.60 \leq r \leq .78$), and large negative relationships with SSIS-TRS Problem Behaviors scores ($-.73 \leq r \leq -.48$).

Discussion

The primary purpose of this study was to examine initial reliability and validity evidence for the ASF. Confirmatory factor analysis indicated that the ASF retains the structure of the original ACES-TF. As predicted, all ASF scores produced reliability coefficients sufficient for individual decision-making (Salvia et al., 2010). With regard to convergent validity, the magnitude, direction, and pattern of ASF concurrent validity relationships were generally consistent with hypotheses for STAR and SSIS-TRS scores. For example, as expected due to the overlap in constructs, the ASF Interpersonal Skills scale scores demonstrated stronger relationships with SSIS-TRS scale (both Social Skills and Problem Behaviors scales) scores than other ASF scale scores.

One expected pattern did not emerge, however. Specifically, when considering measurement error, the relationships between all three ASF Academic Skills scales and STAR measures were roughly equivalent. This finding indicates that although ASF Academic Skills scale scores appear to measure broad academic skills, these scores may not be specific enough to sufficiently represent their target subdomains. Two other findings underscore this possibility. First, reliability coefficients for ASF Academic Skills scales were so high as to suggest they are measuring the same construct. Examining item content of the ASF Academic Skills scales indicates that such a possibility may not be due to redundancy per se, but rather because some items (e.g., written communication) are dependent on others (e.g., spelling, grammar). Second, interfactor correlations between ASF Academic Skills constructs were very high and uniformly

higher than intercorrelations between ASF Academic Enablers constructs.

From a practical perspective, how problematic these findings are considered depends on the context of measurement. Specifically, practitioners may be willing to sacrifice the "edges" of conceptual construct space to focus on the "core" of the construct of interest and efficiently measure that core. This possibility is especially relevant for situations in which measurement is focused more on identifying students at risk of difficulties rather than providing detailed analysis of strengths and weaknesses. This situation may apply in many measurement contexts focused on academic skills, a domain in which there are a plethora of direct measures available for a variety of different applications such as general outcome measurement (e.g., AIMSweb probes; Pearson, 2012) and comprehensive diagnostic assessment (e.g., WJ-IV Achievement Battery; Schrank, McGrew, & Mather, 2014).

In the academic enablers domain (or noncognitive factors in general), there are far fewer measurement options. Thus, it is encouraging that results support the conclusion that ASF Academic Enabler scales retain the structure of the ACES-TF and are differentially related to validity constructs. Scores from ASF Academic Enablers scales would likely be best used in applied or research contexts requiring a high number of ratings and would likely minimize time burdens without jeopardizing content and construct validity or reliability. In such applications, the time savings could be substantial. Considering an estimated 15-min ACES-TF completion time (DiPerna & Elliott, 2000) and the fact that the ASF includes roughly 40% of ACES-TF items, the ASF would likely save roughly 9 min per administration. Such time savings would quickly compound in situations requiring several ratings. For example, the current study's sample would have required approximately 45 more hours of teachers' time if the ACES-TF had been

completed instead of the ASF. Such time savings are likely to be valued in research and practice applications.

Pending additional validity studies, the ASF holds promise for several applications. First, the measure might function well as a targeted screening measure administered to students at high risk of academic difficulty. Evidence to support this proposed use would include conditional probability analyses substantiating the predictive validity of ASF scores for relevant criteria. Another potential application would be as a tool to facilitate evaluation of intervention outcomes. Such an application would be analogous to general outcome measurement for represented ASF domains similar to brief behavior rating scales developed for social domains (e.g., Gresham et al., 2010). Given the difficulties inherent in measuring change (Cronbach & Furby, 1970) that are especially problematic for rating scales (Hobart, Cano, Zajicek, & Thompson, 2007), further research could focus on developing IRT based scoring procedures to more appropriately assess growth for such an application.

There are several important limitations to consider relative to this study. First, although somewhat racially diverse, the current sample was not representative of the current United States population of children (U.S. Department of Education Office for Civil Rights, 2016). The current sample also included a greater percentage of students from younger grades. Furthermore, although the sample was sufficient for correlational analyses, it was minimally sufficient for confirmatory factor analyses (Kline, 2011). Future research should examine the performance of the ASF with a larger and more diverse sample. Finally, the interval between collection of ASF data and validity measures data was longer than is ideal for examining concurrent relationships.

There are many potential avenues for future research resulting from this study. First, future research should continue to examine ASF scores to ensure they have sufficient reliability and validity evidence to justify their use in research and practice. Future research should also supplement the convergent validity evidence collected as part of this study. Particularly, important construct relationships to examine include convergent correlations with measures assessing similar constructs (e.g., scores from the Learning Behaviors Scale; McDermott et al., 1999) and discriminant validity evidence. Another important future research direction is examining predictive validity, which is especially relevant for screening applications. Similarly, Receiver Operating Characteristic (ROC) curve analysis and conditional probability analysis would be particularly useful for establishing and evaluating screening cut points. Finally, given the indications that ASF Academic Skills scales may not sufficiently differentiate their target constructs, introduction of a limited number (1–2 items per scale) of specific reading, mathematics, or critical thinking items on the “edges”

of construct space may improve the psychometric properties of these scales.

Overall, there is evidence that the ASF generally produces reliable and valid scores while retaining a factor structure consistent with the model of the original ACES-TF. As such, the current study provides evidence for the psychometric adequacy of scores from the ASF that is uncommon in the short form development literature (Smith et al., 2000). Based on studies to date, the ASF holds promise as a brief yet technically sound tool for the examination of several non-cognitive factors. Given recent questions surrounding the adequate measurement of noncognitive factors (Credé et al., 2012; Duckworth & Yeager, 2015), further development and validation efforts such as those in this study will be necessary to promote greater understanding of these constructs and their contributions to learning in schools.

Funding This study was funded by the Institute of Education Sciences (Grant Numbers R305A090438 and R305B090007).

Compliance with Ethical Standards

Conflict of interest James DiPerna is the lead author of the Academic Competence Evaluation Scales.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent and assent was obtained for all participants included in the study.

References

- Anthony, C. J., & DiPerna, J. C. (2017). Identifying sets of maximally efficient items from the Academic Competence Evaluation Scales—Teacher Form. *School Psychology Quarterly*, 32(4), 552.
- Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, 43(4), 972–1059.
- Brady, C. E., Evans, S. W., Berlin, K., Bunford, N., & Kerns, L. (2012). Evaluating school impairment with adolescents using the classroom performance survey. *School Psychology Review*, 41, 429–446.
- Cleary, T. J., Gubi, A., & Prescott, M. V. (2010). Motivation and self-regulation assessments: Professional practices and needs of school psychologists. *Psychology in the Schools*, 47, 985–1002. <https://doi.org/10.1002/pits>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.
- Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology*, 102(4), 874–888. <https://doi.org/10.1037/a0027403>.
- Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of*

- Personality and Social Psychology*, 113(3), 492–511. <https://doi.org/10.1037/pspp0000102>.
- Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological Bulletin*, 74(1), 68.
- Demaray, M. K., & Jenkins, L. N. (2011). Relations among academic enablers and academic achievement in children with and without high levels of parent-rated symptoms of inattention, impulsivity, and hyperactivity. *Psychology in the Schools*, 48, 573–586. <https://doi.org/10.1002/pits.20578>.
- DiPerna, J. C., & Elliott, S. N. (2000). *Academic Competence Evaluation Scales*. San Antonio, TX: The Psychological Corporation.
- Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237–251. <https://doi.org/10.3102/0013189X15584327>.
- Elliott, S. N., & Gresham, F. M. (2007). *Social Skills Improvement System: Classwide Intervention Program guide*. Bloomington, MN: Pearson Assessments.
- Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., et al. (2012). *Teaching adolescents to become learners. The role of noncognitive factors in shaping school performance: A critical literature review*. Chicago: University of Chicago Consortium on Chicago School Research.
- Field, A. (2009). *Discovering statistics using SPSS*. Sage publications.
- Garcia, E. (2014). *The need to address noncognitive skills in the education policy agenda (Briefing Paper No. 386)*. Retrieved November 14, 2017 from Economic Policy Institute. <http://files.eric.ed.gov/fulltext/ED558126.pdf>.
- Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System (SSRS)*. Circle Pines, MN: American Guidance Service.
- Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System-Rating Scales*. Minneapolis, MN: Pearson Assessments.
- Gresham, F. M., Elliott, S. N., Cook, C. R., Vance, M. J., & Kettler, R. (2010). Cross-Informant agreement for ratings for social skill and problem behavior ratings: An investigation of the Social Skills Improvement System-Rating Scales. *Psychological Assessment*, 22, 157–166. <https://doi.org/10.1037/a0018124>.
- Gresham, F. M., Elliott, S. N., Vance, M. J., & Cook, C. R. (2011). Comparability of Social Skills Rating System to the Social Skills Improvement System: Content and psychometric comparisons across elementary and secondary age levels. *School Psychology Quarterly*, 26, 27–44. <https://doi.org/10.1037/a0022662>.
- Hambleton, R. K. (2010). Review of the Academic Competence Evaluation Scales. In R. A. Spies, J. F. Carlson, & K. F. Geisinger (Eds.), *The eighteenth mental measurements yearbook* (pp. 1–4). Lincoln, NE: Burns Institute of Mental Measurements.
- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411–482.
- Hobart, J. C., Cano, S. J., Zajicek, J. P., & Thompson, A. J. (2007). Rating scales as outcome measures for clinical trials in neurology: Problems, solutions, and recommendations. *The Lancet Neurology*, 6(12), 1094–1105.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. <https://doi.org/10.1080/10705519909540118>.
- Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success (No. w20749)*. National Bureau of Economic Research.
- Kline, R. B. (2011). *Principles and practices of structural equation modeling*. New York, NY: Guilford.
- Malecki, C. K., & Elliot, S. N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly*, 17, 1–23.
- McCormick, M. P., O'Connor, E. E., Cappella, E., & McClowry, S. G. (2013). Teacher-child relationships and academic achievement: A multilevel propensity score model approach. *Journal of School Psychology*, 51, 611–624. <https://doi.org/10.1016/j.jsp.2013.05.001>.
- McDermott, P. A., Green, L. F., Francis, J. M., & Stott, D. H. (1999). *Learning Behaviors Scale*. Philadelphia: Edumetric and Clinical Science. <https://doi.org/10.1521/scpq.17.1.1.19902>.
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (7th Ed.). Los Angeles, CA: Muthén & Muthén.
- Pearson. (2012). *AIMSweb technical manual*. Bloomington, MN: Pearson. Retrieved from www.aimsweb.com/wp-content/uploads/aimsweb-Technical-Manual.pdf.
- Renaissance Learning. (2012). *STAR Math technical manual*. Wisconsin Rapids, WI: Renaissance Learning.
- Renaissance Learning. (2015). *STAR Reading technical manual*. Wisconsin Rapids, WI: Renaissance Learning.
- Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children* (2nd ed.). Circle Pines, MN: American Guidance Service.
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–373. <https://doi.org/10.1037/a0029315>.
- Rosen, J. A., Glennie, E. J., Dalton, B. W., Lennon, J. M., & Bozick, R. N. (2010). *Noncognitive skills in the classroom: New perspectives on educational research*. Research Triangle Park, NC: RTI International.
- Sabers, D. L., & Bonner, S. (2010). Review of the Academic Competence Evaluation Scales. In R. A. Spies, J. F. Carlson, & K. F. Geisinger (Eds.), *The eighteenth mental measurements yearbook* (pp. 4–6). Lincoln, NE: Burns Institute of Mental Measurements.
- Salvia, J., Ysselydke, J. E., & Bolt, S. (2010). *Assessment in special and inclusive education* (11th ed.). Boston: Houghton Mifflin.
- Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV tests of achievement*. Rolling Meadows, IL: Riverside.
- Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, 12(1), 102–111. <https://doi.org/10.1037/1040-3590.12.1.102>.
- U.S. Department of Education Office for Civil Rights. (2016). *Civil rights data collection: A first look*. Retrieved December 12, 2017 from <https://www2.ed.gov/about/offices/list/ocr/docs/2013-14-first-look.pdf>.
- Volpe, R. J., DuPaul, G. J., DiPerna, J. C., Jitendra, A. K., Lutz, G. L., Tresco, K., et al. (2006). Attention deficit hyperactivity disorder and scholastic achievement: A model of mediation via academic enablers. *School Psychology Review*, 35, 47–61.
- Wechsler, D. (2002). *Wechsler individual achievement test* (2nd ed.). San Antonio, TX: Psychological Corp.
- West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F., et al. (2016). Promise and Paradox-measuring students' non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis*, 38, 148–170. <https://doi.org/10.3102/0162373715597298>.
- Zaslow, M., Halle, T., Martin, L., Cabrera, N., Calkins, J., Pitzer, L., et al. (2006). Child outcome measures in the study of child care quality. *Evaluation Review*, 30, 577–610. <https://doi.org/10.1177/0193841X06291529>.
- Zeehandelaar, D., & Winkler, A. M. (2013). *What parents want: Education preferences and trade-offs*. Washington, DC: Thomas B. Fordham Institute.