Tools

October 2019

A review of instruments for measuring social and emotional learning skills among secondary school students

Joshua Cox
Brandon Foster
David Bamat
Education Development Center, Inc.

In collaboration with the Social and Emotional Learning Research Alliance

Summary

This resource is intended to support state and local education agencies in identifying reliable and valid instruments to measure three social and emotional learning skills among secondary school students: collaboration, perseverance, and self-regulated learning.

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

REL NORTHEAST & ISLANDS
Regional Educational Laboratory
At Education Development Center, Inc.

U.S. Department of Education
Betsy DeVos, *Secretary*

Institute of Education Sciences
Mark Schneider, *Director*

National Center for Education Evaluation and Regional Assistance
Matthew Soldner, *Commissioner*
Elizabeth Eisner, *Associate Commissioner*
Amy Johnson, *Action Editor*
Felicia Sanders, *Project Officer*

REL 2020–010

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

October 2019

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-17-C-0008 by Regional Educational Laboratory Northeast & Islands administered by Education Development Center, Inc. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Cox, J., Foster, B., & Bamat, D. (2019). *A review of instruments for measuring social and emotional learning skills among secondary school students* (REL 2020–010). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. Retrieved from http://ies.ed.gov/ncee/edlabs.

This report is available on the Regional Educational Laboratory website at http://ies.ed.gov/ncee/edlabs.

Summary

This resource was designed to support state and local education agencies in identifying reliable and valid instruments for measuring collaboration, perseverance, and self-regulated learning among secondary school students. It was developed through the Regional Educational Laboratory Northeast & Islands Social and Emotional Learning Alliance, whose members aim to identify and synthesize emerging evidence on social and emotional learning to improve education outcomes. The alliance's focus on social and emotional learning skills is supported by evidence suggesting that these skills may improve students' academic and career outcomes (Farrington et al., 2012; Gayl, 2017; Heckman, 2008; West et al., 2016). This resource can help alliance members and other state and local education agencies develop and track students' social and emotional learning skills as an indicator of student success within accountability models required by the Every Student Succeeds Act of 2015.

This resource supports stakeholders in:

- Identifying available instruments for measuring collaboration, perseverance, and self-regulated learning among secondary school populations.
- Understanding the information about reliability and validity that is available for each of these instruments.

In addition, the resource offers questions that schools and districts should consider when reviewing reliability and validity information.

Contents

Why this resource?

Students who enhance their social and emotional learning skills may also improve their academic and career outcomes (Farrington et al., 2012; Gayl, 2017; Heckman, 2008; West et al., 2016). These skills may also be malleable and amenable to intervention (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; What Works Clearinghouse, 2007). Accordingly, some education agencies have started to develop and track students' social and emotional learning skills as an indicator of student success within accountability models required by the Every Student Succeeds Act of 2015.

However, researchers have not reached consensus on the best ways to measure social and emotional learning skills or the appropriate uses of existing instruments (see, for example, Duckworth & Yeager, 2015). Some researchers have cautioned against using instruments to measure social and emotional learning skills for summative purposes in high-stakes accountability models because most instruments are fairly new and lack adequate information on their reliability and validity (Duckworth & Yeager, 2015; Melnick, Cook-Harvey, & Darling-Hammond, 2017). Reliability refers to whether an instrument consistently measures the skill across respondents, time, or raters. Validity refers to whether an instrument measures what it intends to measure and whether the inferences drawn from an instrument are appropriate. Indeed, this review of instruments for measuring social and emotional learning skills found no explicit language from their developers suggesting that any of them should be used for summative purposes. It may be more appropriate for schools to use data collected from instruments that measure social and emotional learning skills for formative purposes to inform teaching, learning, and program investments (Melnick et al., 2017). To make these decisions, state and local education agency administrators need information to better understand the intended uses and limitations of available instruments, as well as their reliability and validity.

The Regional Educational Laboratory Northeast & Islands Social and Emotional Learning Alliance is a researcher–practitioner partnership comprising researchers, regional educators, policymakers, and others. Alliance partners share an interest in identifying and synthesizing evidence on social and emotional learning skills to improve education outcomes. Champlain Valley School District (CVSD) in Vermont, one of the alliance's partners, seeks to identify instruments for measuring three social and emotional learning skills described in its mission statement: collaboration, perseverance, and self-regulated learning (Champlain Valley School District, 2016). Its intention is to better understand students' intrapersonal and interpersonal competencies, the two main areas into which social and emotional learning competencies are organized. Intrapersonal competencies refer to how students deal with their own thoughts and emotions, and interpersonal competencies refer to the skills and attitudes students use to interact with others.

CVSD is exploring ways to use scores from instruments that measure collaboration, perseverance, and self-regulated learning to formally evaluate students' social and emotional learning skills, develop its formative and summative assessment systems in accordance with the Every Student Succeeds Act, and provide guidance for educators on measuring and integrating social and emotional learning skills into instruction.

What is this resource?

This resource supports local and state education agencies' efforts to identify reliable and valid instruments that measure social and emotional learning skills (see box 1 for an overview of methods and appendix A for more detailed information). The resource also provides questions that schools and districts should consider when reviewing reliability and validity information. Where possible, the resource describes whether the available reliability and validity information meets conventionally accepted criteria. In recognition that each school or district context is unique, this resource does not recommend specific instruments to use with accountability models required by the Every Student Succeeds Act.

This resource supports stakeholders in:

- Identifying available instruments for measuring collaboration, perseverance, and self-regulated learning among secondary school populations.
- Understanding information about reliability and validity that is available for each of these instruments.

Box 1. Methodology

Instrument identification process

The instruments were identified and reviewed in a six-step process:

1. Described collaboration, perseverance, and self-regulated learning skills.
2. Identified terms related to collaboration, perseverance, and self-regulated learning.
3. Searched the literature for relevant instruments measuring collaboration, perseverance, and self-regulated learning.
4. Determined the eligibility for inclusion in this resource of instruments identified in the literature search.
5. Reviewed the reliability and validity information available for eligible instruments.
6. Determined whether the available reliability and validity information met conventionally accepted criteria.

Construct descriptions

To undertake the first step listed above, the Regional Educational Laboratory Northeast & Islands and Champlain Valley School District (CVSD) partnered to describe each of the social and emotional learning skills (or constructs) to be addressed in this resource:

- *Collaboration:* collaborating effectively and respectfully to enhance the learning environment.
- *Perseverance:* persevering when challenged, dealing with failure in a positive way, and seeking to improve one's performance.
- *Self-regulated learning:* taking initiative and responsibility for learning.

These descriptions align with commonly used definitions in the research (for example, Kuhn, 2015; Shechtman, DeBarger, Dornsife, Rosier, & Yarnall, 2013; Zimmerman, 2000).[1]

Eligibility criteria

Instruments had to meet the following eligibility criteria to be included in this resource:

- Measure one of the three targeted social and emotional learning skills.

(continued)

Box 1. Methodology *(continued)*

- Have been used with a population of secondary school students (students in grades 9–12) in the United States.[2]
- Be publicly available online at no or low cost.[3]
- Be published or had psychometric validation work completed between 1994[4] and 2017.
- Not be published as part of a doctoral dissertation.[5]

As recommended by the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), this resource presents information on fairness and reliability. Six components of validity were also evaluated for each instrument following Messick's (1995) construct validity framework: content, substantive, structural, external, generalizability, and consequential. See box 2 for definitions of reliability, validity, the six components of validity, and fairness. Each eligible instrument was then categorized using the criteria in box 2 (see appendix B for a summary of psychometric evidence for eligible instruments).

Notes

1. There are many definitions of these skills in the research literature. See appendix A for additional citations for each skill.

2. CVSD is specifically interested in identifying reliable and valid instruments for secondary school students.

3. Instruments published in research journals at a relatively low cost (less than $50) were included.

4. A group of educators, researchers, and child advocates met in 1994 to discuss social and emotional learning and coined the term (Durlak, Domitrovich, Weissberg, & Gullotta, 2015).

5. These instruments were excluded because they are not typically analyzed with the same rigor as other published instruments (for example, instruments published in peer-reviewed journals).

Source: American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014) and Messick (1995).

This resource indicates whether psychometric information was available for reliability and seven components of validity—content, structural, external, consequential, generalizability, fairness, and substantive[1] (see box 2 for definitions). Schools and districts can use reliability and validity information to evaluate whether the adoption of a measure will likely produce sufficient information to meet the needs of that school or district. It is important to note that the mere availability of reliability and validity information for a measure does not necessarily indicate support for the use of that measure.

Box 2. Definitions of key terms

Reliability. Whether an instrument consistently measures the skill across respondents, time, or raters. Information that could support reliability includes several families of statistics, including measures of internal consistency such as Cronbach's α, omega or test-retest associations, and inter-rater reliability. For the instruments included in this resource, evaluation of reliability information was based on the widely cited conventional criterion of a Cronbach's $\alpha \geq .70$ (Nunnally, 1978).

Validity. Whether an instrument measures what it is intended to measure and whether the inferences drawn from an instrument are appropriate. This resource focuses on the following seven components of validity.

(continued)

Box 2. Definitions of key terms *(continued)*

Content validity. Whether items for an instrument contain content that adequately describes the skill(s). Information that could support this component of validity includes a review of survey items by experts, use of a well constructed theory for the skill(s), and the use of multiple perspectives in creating the items. An example of using multiple perspectives is asking teachers and students to define the skill and using that information with experts to construct items.

Structural validity. Whether an instrument's items relate statistically to the skill being measured. Information that could support this component of validity includes advanced statistical analyses, typically factor analyses that examine whether items are correlated in expected ways. For example, one approach to measuring structural validity is to show that an instrument's items statistically relate to the skills and associated subscales that the instrument measures.

External validity. Whether there are correlations between scores from the instrument and scores from other instruments measuring similar skills. Information that could support this component of validity includes positive correlations between scores generated from the instrument and other similar instruments measuring the skill. For example, one approach to measuring external validity is to establish whether correlations exist between two different instruments that measure perseverance.

Consequential validity. Whether scores generated from an instrument are associated with the intended consequences of using the instrument, such as improving student outcomes.[1] Information that could support this component of validity includes correlations between students' scores on the measure and grade point average or standardized test scores, if the intended consequence is improved student learning, or graduation or job attainment.

Generalizability validity. Whether scores from the instrument are correlated with other modes of measurement for the same skill, such as self-reported or observational information. Unlike external validity, which explores correlations between scores from different instruments measuring similar skills, generalizability explores correlations between scores from different modes of measurement of the same skill. Information that could support generalizability includes correlations between students' self-report scores and either teacher-report scores for students or other observational instruments measuring student behavior.

Fairness. Whether an instrument is not biased against specific subgroups of students. Information that could support this component of validity includes having versions of the measure available in multiple languages or statistical tests showing that scores from the measure function similarly across all subgroups. For example, this could include students from different racial/ethnic backgrounds, students eligible for the national school lunch program, or English learner students.

Substantive validity. Whether students process an instrument's items or tasks in the way the developers intended. Information that could support this component of validity includes cognitive interviews with a series of students as they engage with items from the measure. For example, interviewing students can provide information about sources of confusion that might emerge as students respond to items on the measure.

Note

1. According to Messick (1995, p. 11), "the consequential aspect of construct validity includes evidence and rationales for evaluating the intended and unintended consequences of score interpretation and use in both the short and long term." In practice, the consequences of using a measure can be quite broad and depend on how a district might propose to use a measure. However, the purpose of this resource is to review instruments that might be predictive of important student outcomes. Thus, a consequence of using one of the instruments in this resource should be that it helps districts focus on developing behaviors or skills that could predict important student outcomes (for example, test scores, graduation, and college enrollment). For this reason the resource defines consequential validity by whether scores generated on the measure are associated with important student outcomes. There are many definitions of these skills in the research literature. See appendix A for additional citations for each skill.

Practitioners using this resource should evaluate whether the reliability and validity information for each instrument provides sufficient support for adopting the measure in their district. Practitioners looking for support in evaluating the psychometric information for each instrument can find a series of guiding questions in worksheet 1 to help identify the information necessary to select instruments that are likely to be useful in various local contexts.

Different uses of an instrument will place unique demands on the quality of reliability and validity information presented. For example, a district could be interested in monitoring perseverance among students. In that case, whether an instrument has demonstrated correlations with other student outcomes (that is, consequential or generalizability validity) may not be a concern. On the other hand, practitioners might be interested in ensuring that an instrument has demonstrated correlations with other student outcomes if their goal is to provide formative feedback to support a broader goal such as improving students' grades or increasing educational attainment. Thus, practitioners should first identify the components of reliability and validity that are most important to their school or district, which puts them in a better position to evaluate reliability and validity information for instruments measuring social and emotional learning skills.

Worksheet 1. Questions to identify and evaluate instruments for measuring social and emotional learning skills

Practitioners can use the questions in this worksheet to identify instruments and determine whether those instruments align with their needs. The questions in step 1 can be used to identify the skills that will be measured, the target group of respondents, and the purpose for using the instrument. The questions in step 2 can be used to identify the components of reliability and validity that are most important in the practitioner's school or district. Practitioners can use their yes/no responses in step 2 and accompanying considerations when reviewing the psychometric information for each instrument presented in appendix B. The worksheet concludes by offering additional considerations to support practitioners in identifying whether administering an instrument is feasible in their school or district.

Step 1. Indicate your response for each of the following questions. Be as specific as possible. Use this information to identify an initial list of instruments that your school or district might use.

What are the specific skills to be measured? That is, are you are interested in measuring collaboration, perseverance, or self-regulated learning? (Then, look at table 1 to see which instruments cover the specific skills of interest to you.)

What students are you planning to assess (for example, all high school students)?

What is the purpose for using an instrument (for example, to provide support to teachers)?

Step 2. Now that you have identified the specific skills to be measured, the target group of respondents, and the purpose for using the instrument, work through the following yes/no questions and associated considerations.

You can use responses to this worksheet to synthesize and evaluate information in the tables in appendix B. You can also use table 2 to quickly identify whether information on reliability and validity is available for each instrument.

Note: For each instrument, practitioners should, at a minimum, consider information presented in appendix B on reliability, content validity, and structural validity. Table 3 in the main text provides information on whether reliability information and structural validity information met conventionally accepted criteria. If this information is not provided, or is provided but is not within conventionally accepted criteria, then scores generated from the instrument may not be useful. Finally, reviewing content validity is necessary since it considers how the content from an instrument's items overlaps with the practitioner's understanding of a particular social and emotional learning skill. The alignment between the developer's definition of a social and emotional learning skill and the practitioner's is a core issue for ensuring that a school or district is measuring what it intends to measure.

Question	Response and considerations
Is it important that students were involved in the instrument's development process?	☐ Yes ☐ No If yes, consider that this resource found no information on **substantive validity** for any of the instruments reviewed.
Are you interested in using scores from the instrument along with instruments that measure other related social and emotional learning skills?	☐ Yes ☐ No If yes, consider examining information presented in appendix B for **external validity** to check whether scores generated from the instrument are related to scores from other conceptually similar instruments of social and emotional learning skills.
Are you interested in measuring a specific social and emotional learning skill using more than one mode of measurement, such as student self-report surveys and observations?	☐ Yes ☐ No If yes, consider examining information presented in appendix B for **generalizability validity** to see if analyses were undertaken to establish whether scores from the instrument are correlated with other modes of measurement of the same social and emotional learning skill.
Are you interested in connecting your students' social and emotional learning skills scores to other consequential outcomes, such as achievement scores, graduation rates, and attendance?	☐ Yes ☐ No If yes, consider information presented in appendix B for **consequential validity** and table 4 in the main text. Specifically, examine whether scores from the instrument are correlated with other desired student outcomes.
Are you interested in comparing scores on the instrument for different subgroups of students (for example, by race/ethnicity, eligibility for the national school lunch program, or English learner student status)?	☐ Yes ☐ No If yes, consider examining information presented in appendix B for **fairness** to see whether information is available about the specific subgroups you are comparing.

Practitioners should also consider the following questions for all instruments under consideration:

- How affordable is the instrument to administer to the target group of respondents? Practitioners should consider costs associated with purchasing the instrument, if applicable, as well as any administrative costs associated with administering and scoring the assessment and reporting the results.
- How much time is available for students to complete an instrument? Practitioners might consider piloting the instrument with a few students to better understand the amount of time that students require to complete the instrument.
- What format of administration is feasible in your context? Practitioners might consider piloting the instrument with a few students to better understand whether the format is feasible in their context.
- How will scores be reported? Are the scores easily interpreted and useful? Practitioners should consider their purpose for administering the instrument (see step 1) and whether the results from the instrument will support that purpose.

What instruments are available?

In total, 16 instruments were assessed to be eligible for inclusion in the resource. The initial search yielded 67 instruments as possible measures of one or more of the social and emotional learning skills of interest. Of those 67 instruments, 30 were excluded because they had not been administered with secondary school students in the United States, 24 were excluded because they were not publicly available, and 7 were excluded because they were not published or had not undergone validation work between 1994 and 2017.[2]

Eligible instruments included five measures of collaboration, four measures of perseverance, four measures of self-regulated learning, and three measures of both perseverance and self-regulated learning (table 1). All instruments measuring perseverance or self-regulated learning were student self-report surveys. Three of the five instruments measuring collaboration were student self-report surveys, one was task or performance based, and one was a teacher-report survey.

Note that different instruments measuring a social and emotional learning skill can define that skill differently. For example, the Motivated Strategies for Learning Questionnaire measures self-regulated learning, which the authors describe as including planning, monitoring, and regulating activities (Pintrich, Smith, García, & McKeachie, 1991, p. 23). The Junior Metacognitive Awareness Inventory, another measure of self-regulated learning, uses a framework focused on planning, monitoring, and evaluation (Sperling, Howard, Miller, & Murphy, 2002, p. 55). These small differences are driven by underlying differences in the theoretical research used to shape the content of items and can ultimately lead to variation between different instruments that measure the skill. Before using any instrument, practitioners should examine items from the instrument to see whether they aligns with their perspective on the skill. Social and emotional skills described in research may share a common name, but they can be defined slightly differently by the items in an instrument.

Instruments differ in their intended purpose, including whether for research, formative, or summative uses. Descriptions for each of these uses is as follows:

- *Research use.* The intention is to use results produced by the instrument to describe these skills for a particular population or to examine relationships. The research may then be used as evidence to inform policy or practice, but it is not typically linked to a punitive or beneficial consequence for individuals, schools, or districts.
- *Formative use.* The intention is to use results produced by the instruments to inform instructional change that can influence positive change in students.
- *Summative use.* The intention is to assign a final rating or score to each student by comparing each student against a standard or benchmark. These comparisons can be used to assess the effectiveness of a teacher, school, or district, and an assessment of underachievement can lead to negative consequences for teachers, schools, or districts. Additionally, these comparisons can be used to determine whether a student should be promoted to the next grade level or graduate. For that reason, summative instruments are perceived as having higher stakes, and instruments used for summative purposes traditionally require more stringent reliability and validity evidence before use (Crooks, Kane, & Cohen, 1996; Haladyna & Downing, 2004).

Table 1. Format of eligible instruments, by social and emotional learning skill

Social and emotional learning skill measured and instrument	Format
Collaboration	
Revised Self-Report Teamwork Scale	Student self-report survey
Teamwork Scale	Student self-report survey
Teamwork Scale for Youth	Student self-report survey
Subjective Judgement Test	Task or performance based
Teacher-Report Teamwork Assessment	Teacher-report survey
Perseverance	
Engagement with Instructional Activity	Student self-report survey
Expectancy-Value-Cost Scale	Student self-report survey
Grit Scale—Original Form	Student self-report survey
Grit Scale—Short Form	Student self-report survey
Self-regulated learning	
Inventory of Metacognitive Self-Regulation on Problem-Solving	Student self-report survey
Junior Metacognitive Awareness Inventory	Student self-report survey
Self-Directed Learning Inventory	Student self-report survey
Self-Regulation Strategy Inventory—Self-Report	Student self-report survey
Perseverance and self-regulated learning	
Motivated Strategies for Learning Questionnaire	Student self-report survey
Program for International Student Assessment Student Learner Characteristics as Learners	Student self-report survey
Student Engagement Instrument	Student self-report survey

Note: Instruments are sorted alphabetically first by the measured skill and second by the format of the instrument.

Source: Authors' analysis based on sources shown in tables B1–B16 in appendix B.

Of the 16 instruments reviewed for this resource, 11 were used for research purposes and 5 as a formative tool. There was no explicit language suggesting that any of the instruments should be used for summative purposes. Practitioners should avoid using an instrument for purposes other than those outlined by the instrument developer unless there is sufficient reliability and validity evidence supporting its use for a different purpose.

What information is available about the reliability of the instruments?

All 16 instruments eligible for inclusion in this resource have information on reliability (table 2; see appendix B for more detailed information). Fifteen of the instruments have information on a measure of reliability known as Cronbach's α, which is used to gauge internal consistency by measuring the extent of the relationship among the items (for example, survey questions) in a measure. Reliability statistics, such as Cronbach's α, are used to examine the likelihood of an instrument generating similar scores under consistent conditions. Twelve of the instruments met conventionally accepted criteria for Cronbach's α, indicating that the instruments generated scores that were reliable (table 3). Two instruments providing measures for Cronbach's α showed mixed results, with some subscales in the instrument meeting conventionally accepted criteria for reliability and some not. One instrument reported a Cronbach's α value that was below conventionally accepted criteria. Instruments that meet conventionally accepted thresholds for reliability might be more suitable for informing policy and practice.

Table 2. Availability of reliability and validity information in the 16 eligible instruments

Instrument	Reliability	Validity						
		Content	Structural	External	Conse-quential	Generaliz-ability	Fairness	Substan-tive
Revised Self-Report Teamwork Scale	●	●	●	●	●	●	●	○
Subjective Judgement Test	●	●	●	●	●	●	●	○
Expectancy-Value-Cost Scale	●	●	●	●	●	●	●	○
Teacher-Report Teamwork Assessment	●	●	○	●	●	●	○	○
Teamwork Scale for Youth	●	●	●	●	○	●	○	○
Grit Scale—Original Form	●	●	●	●	●	○	○	○
Grit Scale—Short Form	●	●	●	●	●	○	○	○
Student Engagement Instrument	●	●	●	●	●	○	○	○
Self-Directed Learning Inventory	●	●	●	●	●	○	○	○
Teamwork Scale	●	●	●	●	●	○	○	○
Self-Regulation Strategy Inventory—Self-Report	●	●	○	●	●	○	○	○
Junior Metacognitive Awareness Inventory	●	●	●	○	●	○	○	○
Motivated Strategies for Learning Questionnaire	●	●	●	●	○	○	○	○
Program for International Student Assessment Student Learner Characteristics as Learners	●	●	●	○	○	○	○	○
Inventory of Metacognitive Self-Regulation on Problem-Solving	●	●	○	○	○	○	○	○
Engagement with Instructional Activity	●	●	○	○	○	○	○	○

● Information is available. ○ Information is not available.

Note: Instruments are sorted according to the availability of reliability and validity information.

Source: Authors' analysis based on sources shown in tables B1–B16 in appendix B.

Table 3. Snapshot of whether reliability and structural validity information met conventionally accepted criteria

Instrument	Information for reliability met conventionally accepted criteria[a]	Information for structural validity met conventionally accepted criteria[b]
Junior Metacognitive Awareness Inventory	Yes	Yes
Program for International Student Assessment Student Learner Characteristics as Learners	Yes	Yes
Self-Directed Learning Inventory	Yes	Yes
Student Engagement Instrument	Yes	Yes
Subjective Judgement Test	Yes	Yes
Teamwork Scale	Yes	Yes
Teamwork Scale for Youth	Yes	Yes
Grit Scale—Original Form	Yes	No
Revised Self-Report Teamwork Scale	Yes	No
Inventory of Metacognitive Self-Regulation on Problem-Solving	Yes	—
Self-Regulation Strategy Inventory—Self-Report	Yes	—
Teacher-Report Teamwork Assessment	Yes	—
Grit Scale—Short Form	Some	Some
Motivated Strategies for Learning Questionnaire	Some	No
Engagement with Instructional Activity	No	—
Expectancy-Value-Cost Scale	—	Yes

Note: Instruments are sorted first according to whether information for the instrument met conventionally accepted criteria for reliability and then for structural validity. The two statistics were evaluated using only the source articles presented for each measure in appendix B. Additional details regarding conventionally accepted criteria are available in appendix A.

a. Evaluation of reliability information was based on the widely cited conventional criterion of a Cronbach's $\alpha \geq .70$ (Nunnally, 1978). However, researchers have highlighted that high Cronbach's α values also correspond with measuring a decreased range of the measured skill (Sijtsma, 2009). *Yes* indicates Cronbach's $\alpha \geq .70$; *No* indicates Cronbach's $\alpha < .70$; *Some* indicates mixed results, with some subscales containing Cronbach's α values that meet the .70 threshold and some containing values that do not; and — indicates that articles did not provide information for Cronbach's α.

b. Evaluation of structural validity information was based on whether the articles reported confirmatory factor models with fit statistics for the models falling into conventionally acceptable ranges: Tucker Lewis index ≥ .90, comparative fit index ≥ .90, standardized root mean square residual ≥ .08, and root mean square error of approximation > .05. *Some* indicates mixed results, with some subscales meeting conventionally accepted criteria while others did not, and — indicates that articles that did not provide information for confirmatory factor tests.

Source: Authors' analysis based on sources shown in tables B1–B16 in appendix B.

What information is available about the validity of the instruments?

All 16 instruments eligible for inclusion in this resource had information related to at least one component of validity (see table 2).

Content validity

The component of validity most commonly available for eligible instruments was content validity (see table 2). While content validity is often supported by reviews involving multiple stakeholders with content expertise, none of the instruments included in this resource had evidence of these types of review. All 16 instruments did, however, offer information

on content validity in discussions about the theory and previous research that guided the construction of items that make up the instruments.

Structural validity

Twelve of the 16 instruments in this resource had information on structural validity (see table 2). Eight of these instruments were supported by evidence indicating that the instrument met conventionally accepted criteria based on results from confirmatory factor analysis, a statistical technique to establish whether the items from a measure are related to the centrally measured skill (see table 3). One additional measure showed mixed results because the authors reported only one of many statistics needed to assess whether the model fit adequately.

The predominant form of information provided for structural validity was results from confirmatory factor analysis. Social and emotional learning skills are often made up of more than one component. For this reason, instruments that measure social and emotional learning skills often comprise subscales that measure the components that make up the broader social and emotional learning skill. For example, developers of the Grit Scale (Duckworth, Peterson, Matthews, & Kelly, 2007) hypothesize that the instrument consists of two separate but related components, Consistency of Interest and Perseverance of Effort. A confirmatory factor analysis of the Grit Scale was conducted to answer whether the data confirmed the existence of these two separate components. All items for a measure should conceptually align with the overall concept of the skill in some way, with each item describing a different component of the skill.

External validity

Twelve of the 16 instruments in this resource had information on external validity (see table 2). External validity refers to information about the extent to which results produced from an instrument are correlated with results from other instruments in strength and direction in the theoretically expected manner. For instance, if instruments for depression, anxiety, and happiness were administered to the same sample, scores on the depression instrument would be expected to be positively associated with scores on the anxiety instrument, while scores on the happiness instrument would be expected to be negatively associated with scores on the depression instrument. If these relationships were not observed, then the instrument likely did not measure what it was intended to measure and its external validity is questionable.

Consequential validity

Eleven of the 16 instruments in this resource had information on consequential validity (see table 2). Consequential validity refers to information about the extent to which results produced by an instrument are associated with important student outcomes. The predominant student outcomes examined in these studies were student achievement outcomes. These associations were in the expected positive direction for 10 of these 11 instruments (table 4). One measure, the Student Engagement Instrument, was also negatively correlated with student suspension rates.

Table 4. Snapshot of consequential validity information

Instrument	Positively correlated with consequential student outcomes[a]	Type of consequential student outcome reported in article
Student Engagement Instrument	Yes	Grade point averages, course grades, and suspension rates
Revised Self-Report Teamwork Scale	Yes	Course grades
Teamwork Scale	Yes	Grade point averages
Expectancy-Value-Cost Scale	Yes	Achievement scores
Junior Metacognitive Awareness Inventory	Yes	Swanson Metacognitive Questionnaire, course grades, and grade point averages
Self-Directed Learning Inventory	Yes	Grade point averages
Self-Regulation Strategy Inventory—Self-Report	Yes	Course grades
Subjective Judgement Test	Yes	Course grades[b]
Teacher-Report Teamwork Assessment	Yes	Course grades
Grit Scale—Short Form	Yes	Achievement scores
Grit Scale—Original Form	No	Achievement scores
Engagement with Instructional Activity	—	—
Inventory of Metacognitive Self-Regulation on Problem-Solving	—	—
Motivated Strategies for Learning Questionnaire	—	—
Program for International Student Assessment Student Learner Characteristics as Learners	—	—
Teamwork Scale for Youth	—	—

Note: Instruments are sorted in the table according to whether they are positively correlated with consequential student outcomes.

a. Yes indicates a positive association with consequential student outcome; no indicates a negative association with consequential student outcome; and — indicates that the articles did not provide information.

b. Although correlations between scores on the Subjective Judgement Test and course grades were in the expected direction, the correlation was not statistically significant.

Source: Authors' analysis based on sources shown in tables B1–B16 in appendix B.

Generalizability, fairness, and substantive validity

Five of the 16 eligible instruments had information on generalizability (see table 2). Generalizability is the extent to which results measuring a trait are associated with results measuring the same trait using a different mode of measurement, such as student self-report and teacher report.

Three of the 16 instruments had information on fairness. Although some of the instruments indicated that the sought-after information existed in multiple languages, the search did not identify information on whether the instruments behave similarly for different social groups.

No instruments had information on substantive validity. Substantive validity refers to whether respondents process items or tasks in the way the developers intended. One way that instrument developers can determine whether there is evidence of substantive validity is to conduct cognitive interviews during instrument development and collect verbal information from individuals as they respond to individual items within a particular measure.

Information collected in these interviews is used to ensure that respondents are engaging with the instrument as the developers intended.

Implications

This resource included 16 publicly available instruments that focus on measuring collaboration, perseverance, and self-regulated learning among secondary school students in the United States. More instruments were initially identified but ultimately excluded because they were not intended for use with secondary school students. Practitioners should use caution when administering any instrument that was not developed for the population of students that the instrument will be used to evaluate because those instruments often lack psychometric evidence for that population. In the absence of such psychometric evidence, practitioners cannot ensure that the analyses of scores generated from the measure are reliable or valid for the target population. For example, analyzing student change across time, comparing key subgroups, and benchmarking current levels of a trait in the district are not warranted.

Among the 16 instruments identified, 11 were developed for use in research and 5 for formative instruction. None of the information collected suggested that the instruments should be used for summative purposes. With schools and districts ramping up efforts to measure social and emotional learning skills for formative and summative use, practitioners would benefit from the development of additional instruments for these purposes. Likewise, additional work is needed to better understand whether existing instruments that were not specifically developed for formative or summative purposes can be used for those purposes. Meanwhile, practitioners should be cautious when using any measure for summative purposes that has not been developed and validated for that purpose. Without evidence to support that an instrument is valid and reliable for a specific purpose, administrators are at risk of using an invalid and unreliable assessment to inform high-stakes decisionmaking.

Finally, none of the instruments identified in this resource had information for substantive validity, and only three had information on fairness. Information for the substantive component of validity is necessary to facilitate understanding of whether respondents process the content of items from a measure as the developers intended. Information on fairness is necessary for evaluating whether the measure is valid for comparing scores between subgroups of students. To help practitioners better understand whether instruments are measuring social and emotional learning skills for all populations of students, instrument developers could assess the substantive and fairness components of validity. Practitioners should use caution when administering instruments that lack information on substantive validity or fairness, since these instruments may not be appropriate for all students that are being evaluated.

Limitations

The first limitation of this resource is that only a small subset of instruments was reviewed. The criteria for inclusion in this resource were more specific than in prior reviews of instruments measuring social and emotional learning skills. For example, only instruments measuring collaboration, perseverance, or self-regulated learning were reviewed. Further, this resource excludes any instrument that had not been used with a secondary school student population in the United States. These criteria were established because CVSD

was interested in identifying instruments that measure collaboration, perseverance, and self-regulated learning in students in secondary school. Similarly, with a focus on identifying instruments that practitioners can use in their schools and districts, this resource excludes any instrument that was not publicly available.

Second, an implicit assumption of the study was that only instruments or validation studies that appeared in response to the search queries were included. It is possible that other instruments or validation studies exist that were not identified in the queries. For example, if a validation study did not include any of the search terms described in appendix A, the psychometric information for that instrument does not appear in this resource.

Appendix A. Methodology

The review of instruments included six primary steps:

- Described collaboration, perseverance, and self-regulated learning skills.
- Identified terms related to collaboration, perseverance, and self-regulated learning.
- Searched the literature for relevant instruments measuring collaboration, perseverance, and self-regulated learning.
- Determined the eligibility for inclusion in this resource of instruments identified in the literature search.
- Reviewed the reliability and validity information available for eligible instruments.
- Determined whether the reliability and validity information met conventionally accepted criteria.

Described collaboration, perseverance, and self-regulated learning skills

The Regional Educational Laboratory (REL) Northeast & Islands and Champlain Valley School District (CVSD) collaborated to describe each of the social and emotional learning skills (commonly referred to as constructs in the research literature) to be addressed in this resource:

- Collaboration: collaborating effectively and respectfully to enhance the learning environment.
- Perseverance: persevering when challenged, dealing with failure in a positive way, seeking to improve one's performance.
- Self-regulated learning: taking initiative in and responsibility for learning.

These descriptions align with commonly used definitions in the research for collaboration (Johnson & Johnson, 1994; Kuhn, 2015; Wang, MacCann, Zhuang, Liu, & Roberts, 2009); perseverance (Duckworth & Quinn, 2009; Eccles, Wigfield, & Schiefele, 1998; Pintrich, 2003; Seifert, 2004; Shechtman et al., 2013); and self-regulated learning (Muis, Winne, & Jamieson-Noel, 2007; Zimmerman, 2000).

Identified terms related to collaboration, perseverance, and self-regulated learning

The study team identified terms that are synonymous with or related to collaboration, perseverance, and self-regulated learning. For collaboration, related terms included:

- Collaborative competence.
- Collaborative learning.
- Collaborative problem-solving.
- Cooperation.
- Cooperative learning.
- Teamwork.

For perseverance, related terms included:

- Academic courage.
- Academic motivation.
- Conscientiousness.
- Coping.
- Delayed gratification.
- Determination.

- Grit.
- Locus of control.
- Motivation.
- Persistence.
- Resilience.
- Self-control.
- Self-discipline.
- Self-management.
- Tenacity.

For self-regulated learning, related terms included:

- Achievement goals.
- Metacognition.
- Motivation.
- Self-efficacy.
- Task interest.

Searched the literature for relevant instruments

Next, both peer-reviewed academic literature and reports by practitioners and districts were searched. The following were among the main sources searched:

- Academic journal databases such as EBSCOhost (https://www.ebscohost.com) and the Education Resources Information Center (ERIC) (https://eric.ed.gov).
- Institute of Education Sciences publications (https://ies.ed.gov).
- National Science Foundation (https://www.nsf.gov) publications and databases of instruments (for example, through the STEM Learning and Research Center at Education Development Center, http://stelar.edc.org).
- Compendia of instruments and literature reviews (Atkins-Burnett, Fernández, Akers, Jacobson, & Smither-Wulsin, 2012; Fredricks & McColskey, 2012; Fredricks et al., 2011; Kafka, 2016; National Center on Safe Supportive Learning Environments, 2017; O'Conner, De Feyter, Carr, Luo, & Romm, 2017; Rosen, Glennie, Dalton, Lennon, & Bozick, 2010).
- Other relevant databases such as the American Psychological Association PsycTESTS database (http://www.apa.org/pubs/databases/psyctests/) and the Rush University Neurobehavioral Center's SELweb database (http://rnbc.org/research/selweb/).
- Other relevant sources, including those from the following organizations:
  - The Collaborative for Academic, Social, and Emotional Learning at the University of Illinois at Chicago (http://www.casel.org).
  - The National Academies of Sciences, Engineering, and Medicine (http://www.nationalacademies.org).
  - The Ecological Approach to Social Emotional Learning Laboratory at the Harvard Graduate School of Education (https://easel.gse.harvard.edu).
  - The Massachusetts Consortium for Innovative Education Assessment (http://www.mciea.org).
  - P21 Partnership for 21st Century Learning (http://www.p21.org).
  - The Character Lab (https://characterlab.org).

Searches were also conducted for performance assessments, self-report surveys, and other instruments that seek to measure collaboration, perseverance, and self-regulated learning. For example, search terms related to perseverance included:

- [Assessment] AND [Student] AND [Perseverance].
- [Instrument] AND [Student] AND [Perseverance].
- [Performance-based] AND [Assessment] AND [Student] AND [Perseverance].
- [Research] AND [Studies] AND [Student] AND [Perseverance].
- [Survey] AND [Student] AND [Perseverance].

Searches also included terms related to collaboration, perseverance, and self-regulated learning. For example, a secondary search focused on identifying instruments that measure perseverance included:

- [Assessment] AND [Student] AND [Persistence].
- [Instrument] AND [Student] AND [Persistence].
- [Performance-based] AND [Assessment] AND [Student] AND [Persistence].
- [Research] AND [Studies] AND [Student] AND [Persistence].
- [Survey] AND [Student] AND [Persistence].

Determined the eligibility of instruments identified in the literature search

A study team member used a structured protocol to determine the eligibility of instruments identified in the literature search. A second study team member then checked the instruments identified as meeting the eligibility criteria against the protocol a second time. In cases of disagreement on the eligibility of an instrument, the two study team members met to discuss and resolve the discrepancy.

Instruments were deemed eligible if they:

- Measured one of the three targeted social and emotional learning skills.
- Were used with a population of secondary school students in the United States.
- Were publicly available online at no or low cost.
- Were published or had psychometric validation work completed between 1994 and 2017.
- Were not published as part of a doctoral dissertation.

Reviewed the reliability and validity information available for eligible instruments

A second search procedure was carried out for each of the instruments that met the initial screening criteria to identify studies that might provide information about the reliability and validity properties for each instrument. Search terms included: [Name of the instrument] AND [terms that included but were not limited to psychometrics, measurement, reliability, and validity].

The reliability and validity information for eligible instruments was then categorized and summarized to assess whether a measure was reliable and valid to use. Messick's (1995) construct validity framework was used in evaluating six components of validity for each instrument: content, substantive, structural, external, generalizability, and consequential. In addition, this resource presents information on fairness and reliability, as recommended in Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, and National Council on Measurement

in Education, 2014; see box 2 in main text for definitions). Each eligible instrument was then categorized using the criteria defined in box 2 in the main report (see appendix B for a summary of psychometric evidence for eligible instruments).

The reliability and validity evidence available for each eligible instrument was independently reviewed by two study team members. When the evidence identified for an instrument differed, the two study team members met to discuss and resolve the discrepancy.

Independent, external feedback was solicited from two groups: Technical Working Group members, to ensure rigor in methodology and significance in content knowledge, and a project advisory committee, to ensure relevance. Members of the advisory committee included educators and district leaders from CVSD and the Sanborn Regional School District in New Hampshire. The advisory committee collaborated with the study team on research questions, appropriate terminology, analysis, and dissemination strategies.

Determined whether the available reliability and validity information met conventionally accepted criteria

Two components of the psychometric properties of instruments were evaluated to discern whether the information provided met conventionally accepted criteria for optimal performance. These properties were reliability and structural validity.

Evaluation of reliability information was based on the widely cited conventional criterion of a Cronbach's $\alpha \geq .70$ (Nunnally, 1978). However, it should be noted that high Cronbach's α values also correspond with measuring a decreased range of the assessed skill (Sijtsma, 2009).

Most of the information provided for structural validity was in the form of confirmatory factor analyses. For these analyses it is common to provide some indices of model fit. Fit indices include comparative fit index (CFI), root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), and Tucker Lewis Index (TLI). The primary way to evaluate the fit of a model is to examine whether these statistics meet conventionally accepted thresholds. Hu and Bentler (1999) provide a frequently cited framework for evaluating the fit of confirmatory factor models. Conventionally acceptable ranges are CFI > .90, RMSEA > .05, SRMR < .08, and TLI > .90. If the fit statistics provided fell within these thresholds, the measure was considered to have met conventionally accepted criteria for model fit. In some cases, the literature reported combinations of fit statistics, with some that fell within the ranges and some that fell outside them.

Appendix B. Summary of reliability and validity information on eligible instruments

This appendix includes tables summarizing the reliability and validity information identified for each eligible instrument (tables B1–B16). The tables are arranged by the specific social and emotional learning skill they were designed to measure.

Collaboration

Five instruments were intended to measure collaboration (tables B1–B5).

Table B1. Revised Self-Report Teamwork Scale: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Collaboration
Format	Student self-report survey
Number of items	57, but after analyses the authors suggested using just 30
Target population	High school students
Instrument source	Wang, L., MacCann, C., Zhuang, X., Liu, L., & Roberts, R. (2009). Assessing teamwork and collaboration in high school students: A multimethod approach. *Canadian Journal of School Psychology, 24*(2), 41–54.
Past administration of instrument: In school setting?	Yes
Past administration of instrument: In secondary school setting?	Yes
Past administration of instrument: Uses?	Research
Reliability	Cronbach's α: Cooperation (student measure) = .88, Advocate/Guide (student measure) = .80, and Negotiation = .78 (student measure). The conventionally accepted criterion of reliability for Cronbach's α is $\geq$ .70 (Nunnally, 1978).
Content validity	The authors outlined a theory that defines the measured skills.
Substantive validity	Not available
Structural validity	The authors conducted an exploratory factor analysis to examine the dimensional structure of the measures. Three factors emerged: Cooperation, Advocate/Guide, and Negotiation. A confirmatory factor analysis was then used to attempt to confirm that the measure contained three dimensions. This model failed to reach conventionally accepted criteria for good fitting models.
External validity	Between-factor correlations are reported within and across measures.
Generalizability	Correlations are reported for the scores between the measures for the three modes of measurement (student-report, teacher-report, and situational tasks). These are all in the expected direction.
Consequential validity	Correlations are reported between the measures and course grades. The Cooperation factor scores were positively correlated with course grades in science and music; Advocate/Guide scores were positively correlated with course grades in science, social studies, and music; and Negotiation was positively correlated with course grades in math.
Fairness	The authors examined differences in mean scores for demographic subgroups. No significant differences were reported for gender or racial/ethnic subgroups. Significant differences were found for age, where older students scored higher on the instrument.

Source: Wang, L., MacCann, C., Zhuang, X., Liu, L., & Roberts, R. (2009). Assessing teamwork and collaboration in high school students: A multimethod approach. *Canadian Journal of School Psychology, 24*(2), 41–54.

Table B2. Teamwork Scale: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Collaboration
Format	Student self-report survey
Number of items	26
Target population	A high school sample was used for the analyses, but the authors describe the measure as having been developed for college-age students.
Instrument source	French, B., Gotch, C., Immekus, J., & Beaver, J. (2016). An investigation of the psychometric of a measure of teamwork among high school students. *Psychological Test and Assessment Modeling, 58*(3), 455–470.
Past administration of instrument: In school setting?	Yes
Past administration of instrument: In secondary school setting?	Yes
Past administration of instrument: Uses?	Formative
Reliability	Cronbach's α for each subscale ranged from .76 to .92. The conventionally accepted criterion of reliability for Cronbach's α is $\geq$ .70 (Nunnally, 1978).
Content validity	The authors outlined a theory that defines the measured skills.
Substantive validity	Not available
Structural validity	The authors examined nested confirmatory factor analysis models and concluded that the measure exhibited a bifactor model, with a primary factor for Teamwork and subdimensions for Group Composition, Interdependency, Norms and Roles, and Goals. All fit statistics fell within conventionally accepted criteria for good fitting models.
External validity	Correlations between subscales in the measure and academic motivation and group composition are reported. The lowest correlation was r = .47.
Generalizability	Not available
Consequential validity	The Overall Teamwork scale (r = .26) and the Goals subscale (r = .29) were weakly to moderately correlated with grade point average. The Group Composition subscale was moderately correlated with grade point average (r = .35). The Overall Teamwork Scale (r = .26) and the Goals subscale (r = .29) were weakly to moderately correlated with grade point average. The Norms and Roles subscale was moderately correlated with grade point average (r = .35).
Fairness	Not available

Source: French, B., Gotch, C., Immekus, J., & Beaver, J. (2016). An investigation of the psychometric of a measure of teamwork among high school students. *Psychological Test and Assessment Modeling, 58*(3), 455–470.

Table B3. Teamwork Scale for Youth: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Collaboration
Format	Student self-report survey
Number of Items	10, but after analyses the authors suggested removing the first 2 items and using only 8
Target population	The authors describe the sample as "youths." The survey was administered to youths ranging in age from 9 to 15.
Instrument source	Lower, L., Newman, T., & Anderson-Butcher, D. (2015). Validity and reliability of the Teamwork Scale for Youth. *Research on Social Work Practice 27*(6), 1–10.
Past administration of instrument: In school setting?	No, the measure was used in a summer sports-based program.
Past administration of instrument: In secondary school setting?	No, the measure was used in a summer sports-based program.
Past administration of instrument: Uses?	Research
Reliability	Cronbach's α was .79 at pretest, .86 at midpoint, and .88 at post-test. The conventionally accepted criterion of reliability for Cronbach's α is $\geq$ .70 (Nunnally, 1978).
Content validity	The authors noted that researchers with more than 20 years of experience in social work and research methods consulted the literature, examining conceptual and measurement approaches relative to the teamwork and related skills. They used this information to develop the 10 items.
Substantive validity	Not available
Structural validity	The authors hypothesized that the measure contained two factors. Specifically, two items aligned with attitudes toward teamwork, while the remaining items aligned with teamwork behaviors. Confirmatory factor analysis was conducted on the items. Two items were removed because of poor item functioning in the analyses. The hypothesized two-factor structure did not adequately fit the data. A one-factor model fit the data adequately; two problematic items were removed.
External validity	Total scores created for the measure were used in correlational analyses. A significant positive relationship with perceived belonging (r = .41) was indicated, and positive relationships with both social competence (r = .47) and commitment (r = .42) were indicated.
Generalizability	To determine whether the measure could be used longitudinally to evaluate changes in the measured skill over time, the authors examined whether the factor structure was invariant across three time points in the summer program and found "moderate evidence" of invariance across the three time points.
Consequential validity	Not available
Fairness	Not available

Source: Lower, L., Newman, T., & Anderson-Butcher, D. (2015). Validity and reliability of the Teamwork Scale for Youth. *Research on Social Work Practice 27*(6), 1–10.

Table B4. Subjective Judgement Test: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Collaboration
Format	Task or performance based
Number of items	8 situational tasks
Target population	High school students
Instrument source	Zhuang, X., MacCann, C., Wang, L., Liu, O. L., & Roberts, R. D. (2008, October). *Development and validity evidence supporting a teamwork and collaboration assessment for high school students.* Research Report RR-08–50. Ewing, NJ: ETS. https://pdfs.semanticscholar.org/f83e/641f4875466adbed23b353494bc0f6a9d250.pdf.
Past administration of instrument: In school setting?	Yes
Past administration of instrument: In secondary school setting?	Yes
Past administration of instrument: Uses?	Research
Reliability	Cronbach's α = .71. The conventionally accepted criterion of reliability for Cronbach's α is $\geq$ .70 (Nunnally, 1978).
Content validity	The authors outlined a theory that defines the measured skills. In addition, experts were consulted for selecting a scoring mechanism and interpreting results.
Substantive validity	Not available
Structural validity	Exploratory factor analysis showed that the measure contained one factor. This was confirmed through a confirmatory factor analysis that contained fit statistics within the conventionally accepted criteria for good fitting models. Latent class analysis demonstrated that the measure could differentiate between high and low levels of teamwork.
External validity	Authors examined correlations with Myers and Briggs' Big 5 personality test. No significant correlations are reported.
Generalizability	Teachers' self-report scores for the three dimensions of collaboration (collaboration, advocating/influence, and negotiation) were all positively correlated with the Situational Judgement Test tasks and teacher report measures for collaboration in the range of r = .33 to .60.
Consequential validity	The Situational Judgement Test scores did not correlate significantly with course grades (although correlations were in the expected direction).
Fairness	No significant gender differences were found for the three student self-report subscales, teacher-report scores, or Situational Judgement Test scores. In addition, there were no significant differences by racial/ethnic subgroup for any of the measures.

Source: Zhuang, X., MacCann, C., Wang, L., Liu, O. L., & Roberts, R. D. (2008, October). *Development and validity evidence supporting a teamwork and collaboration assessment for high school students.* ETS Research Report RR-08–50. Ewing, NJ: ETS. https://pdfs.semanticscholar.org/f83e/641f4875466adbed23b353494bc0f6a9d250.pdf.

Table B5. Teacher-Report Teamwork Assessment: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Collaboration
Format	Teacher-report survey
Number of items	10
Target population	High school students
Instrument source	Zhuang, X., MacCann, C., Wang, L., Liu, O. L., & Roberts, R. D. (2008, October). *Development and validity evidence supporting a teamwork and collaboration assessment for high school students.* Research Report RR-08–50. Ewing, NJ: ETS. https://pdfs.semanticscholar.org/f83e/641f4875466adbed23b353494bc0f6a9d250.pdf.
Past administration of instrument: In school setting?	Yes
Past administration of instrument: In secondary school setting?	Yes
Past administration of instrument: Uses?	Research
Reliability	Cronbach's α was .98. The conventionally accepted criterion of reliability for Cronbach's α is $\geq$ .70 (Nunnally, 1978).
Content validity	The authors outlined a theory that defines the measured skills.
Substantive validity	Not available
Structural validity	The authors detected a one-factor solution using exploratory factor analysis that represented teacher's self-report of students' teamwork as defined by items pertaining to cooperation, leadership, and conflict resolution. No confirmatory factor analyses were provided for the model.
External validity	The authors examined correlations with the Myers and Briggs Big 5 personality test. No significant correlations were reported.
Generalizability	The authors examined teachers' self-report scores and the three dimensions of collaboration from the student self-report version of the measure, as well as scores from the Situational Judgement Test measure. Positive correlations in the range of r = .14 to r = .33 were reported for all correlations.
Consequential validity	The teacher-report scale correlated significantly with course grades in math (r = .21), science (r = .30), and social studies (r = .27).
Fairness	Not available

Source: Zhuang, X., MacCann, C., Wang, L., Liu, O. L., & Roberts, R. D. (2008, October). *Development and validity evidence supporting a teamwork and collaboration assessment for high school students.* ETS Research Report RR-08–50. Ewing, NJ: ETS. https://pdfs.semanticscholar.org/f83e/641f4875466adbed23b353494bc0f6a9d250.pdf.

Perseverance

Four instruments were designed to measure perseverance (tables B6–B9).

Table B6. Engagement with Instructional Activity: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Perseverance
Format	Student self-report survey
Number of items	4
Target population	Elementary, middle, and high school students
Instrument source	Marks, H. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal, 37*(1), 153–184.
Past administration of instrument: In school setting?	Yes
Past administration of instrument: In secondary school setting?	Not available
Past administration of instrument: Uses?	Research
Reliability	Cronbach's α was .69. The conventionally accepted criterion of reliability for Cronbach's α is $\geq$ .70 (Nunnally, 1978). Pseudo intraclass correlation was .37.
Content validity	The authors outlined a theory that defines the measured skills.
Substantive validity	Not available
Structural validity	Not available
External validity	Not available
Generalizability	Not available
Consequential validity	Not available
Fairness	Not available

Source: Marks, H. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal, 37*(1), 153–184.

Table B7. Expectancy-Value-Cost Scale: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Perseverance
Format	Student self-report survey
Number of Items	10
Target population	Middle and high school students
Instrument source	Kosovich, J., Hulleman, C., Barron, K., & Getty, S. (2014). A practical measure of student motivation: Establishing validity evidence for the Expectancy-Value-Cost Scale in middle school. *Journal of Early Adolescence 27*(1), 1–27.
Past administration of instrument: In school setting?	Yes
Past administration of instrument: In secondary school setting?	Not available
Past administration of instrument: Uses?	Formative
Reliability	Coefficient omega statistics were calculated for each subscale (Expectancy, Value, and Cost) and split by subject area (math and science). For math, omegas were .88 for Expectancy, .84 for Value, and .86 for Cost. For science, omegas were .88 for Expectancy, .88 for Value, and .87 for Cost. Test-retest reliability was provided by subject area (math and science) from fall to winter and ranged from r = .62 to r = .80.
Content validity	The authors outlined a theory that defines the measured skills.
Substantive validity	Not available
Structural validity	The authors conducted extensive confirmatory factor analysis of four different internal conceptualizations of the scale. Authors settled on a three-factor solution with factors for Expectancy, Value, and Cost. The model displayed appropriate fit within conventionally accepted criteria.
External validity	The authors examined relationships between the subscale scores for Expectancy, Value, and Cost. As expected, correlations among the subscales of the measure were more strongly related within subject areas than across them. For example, math Expectancy and math Value were moderately correlated (r = .55), whereas math Expectancy and science Value were less strongly correlated (r = .31). Math Expectancy and science Expectancy were also weakly correlated (r = .29), providing evidence of cross-domain discrimination between measured skills.
Generalizability	The authors examined the longitudinal factor invariance of the three-factor model, finding that latent and observed change were similar and that practitioners could use the observed change as a reliable indicator of changes in measured skills over time.
Consequential validity	Expectancy and Value factor scores were positively correlated with students' achievement scores in math and science in the range of r = .15 to r = .47.
Fairness	Invariance tests demonstrated that the factor structure was similar for boys and girls.

Source: Kosovich, J., Hulleman, C., Barron, K., & Getty, S. (2014). A practical measure of student motivation: Establishing validity evidence for the Expectancy-Value-Cost Scale in middle school. *Journal of Early Adolescence 27*(1), 1–27.

Table B8. Grit Scale—Original Form: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Perseverance
Format	Student self-report survey
Number of items	12
Target population	All ages
Instrument source	Duckworth, A., Peterson, C., Matthews, M., & Kelly, D. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology 92*(6), 1087–1101.
Past administration of instrument: In school setting?	Yes
Past administration of instrument: In secondary school setting?	Not available
Past administration of instrument: Uses?	Research
Reliability	Cronbach's α was .85 for the overall scale, .85 for the Consistency of Interest subscale, and .78 for the Perseverance of Effort subscale. The conventionally accepted criterion of reliability for Cronbach's α is $\geq$ .70 (Nunnally, 1978).
Content validity	The authors outlined a theory that defined the measured skills. In addition, they described shaping the content of the items based on the high-achieving individuals' characteristic attitudes and behaviors that were revealed through initial exploratory interviews with lawyers, business people, academics, and other professionals.
Substantive validity	Not available
Structural validity	The authors first examined item-total correlations, internal reliability coefficients, redundancy, and simplicity of vocabulary to eliminate 10 items from the measure. They followed up with an exploratory factor analysis to understand the dimensional structure of the measure. A two-factor solution emerged, with dimensions for Consistency of Interest and Perseverance of Effort. The exploratory solution for the two dimensions was then tested using a confirmatory factor analysis model. The reported statistics did not meet conventionally accepted criteria.
External validity	The authors examined Pearson's correlation coefficients with the five factors from the Myers and Briggs Big 5 personality test and observed that Grit Scale scores were positively correlated with the Conscientiousness factor (r = .77) and negatively correlated with the Neuroticism factor (r = –.38). Both correlations were in the expected direction.
Generalizability	Not available
Consequential validity	Grit scores were negatively associated with SAT scores for undergraduates (r = –.20). This correlation was not in the expected direction.
Fairness	Not available

Source: Duckworth, A., Peterson, C., Matthews, M., & Kelly, D. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*(6), 1087–1101.

Table B9. Grit Scale—Short Form: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Perseverance
Format	Student self-report survey
Number of items	8
Target population	All ages
Instrument source	Duckworth, A., & Quinn, P. (2009). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment, 91*(2), 166–174.
Past administration of instrument: In school setting?	Yes
Past administration of instrument: In secondary school setting?	Yes
Past administration of instrument: Uses?	Research
Reliability	Cronbach's α values range from .73 to .83 for the overall Grit Scale–Short Form, from .60 to .78 for the Perseverance of Effort subscale, and from .73 to .79 for the Consistency of Interest subscale. The conventionally accepted criterion of reliability for Cronbach's α is $\geq$ .70 (Nunnally, 1978).
Content validity	The authors outlined a theory that defines the measured skills.
Substantive validity	Not available
Structural validity	A confirmatory factor analysis for a model containing factors for Perseverance of Effort and Consistency of Interest showed adequate fit with the one relative fit index provided.
External validity	Factor scores for the two subskills (Perseverance of Effort and Consistency of Interest) were correlated in the expected direction with validation scales in the design. As expected, Grit Scale scores were positively related with the Conscientiousness factor.
Generalizability	Not available
Consequential validity	Results from the Grit Scale were positively and significantly related to student achievement.
Fairness	Not available.

Source: Duckworth, A., & Quinn, P. (2009). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment, 91*(2), 166–174.

Self-regulated learning

Four instruments were designed to measure self-regulated learning (tables B10–B13).

Table B10. Inventory of Metacognitive Self-Regulation on Problem-Solving: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Self-regulated learning
Format	Student self-report survey
Number of items	37
Target population	Middle and high school students
Instrument source	Howard, B., McGee, S. Shia, R., & Hong, N. (2000, April). *Metacognitive self-regulation and problem-solving: Expanding the theory base through factor analysis.* Paper presented at the Annual Meeting of the American Educational Research Association, April 24–28, 2000, New Orleans, LA. https://eric.ed.gov/?id=ED470973.
Past administration of instrument: In school setting?	Yes
Past administration of instrument: In secondary school setting?	Yes
Past administration of instrument: Uses?	Research
Reliability	Cronbach's α was .94 for all items and ranged from .72 to .87 for factors that emerged from the exploratory factor analysis. The conventionally accepted criterion of reliability for Cronbach's α is $\geq$ .70 (Nunnally, 1978).
Content validity	The authors outlined a theory that defines the measured skills.
Substantive validity	Not available
Structural validity	The authors used an exploratory factor analysis, which resulted in five factors. No attempt was made to explain what the factors mean, nor was a confirmatory factor analysis model constructed.
External validity	Not available
Generalizability	Not available
Consequential validity	Not available
Fairness	Not available

Source: Howard, B., McGee, S. Shia, R., & Hong, N. (2000, April). *Metacognitive self-regulation and problem-solving: Expanding the theory base through factor analysis.* Paper presented at the Annual Meeting of the American Educational Research Association, April 24–28, 2000, New Orleans, LA. https://eric.ed.gov/?id=ED470973.

Table B11. Junior Metacognitive Awareness Inventory: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Self-regulated learning
Format	Student self-report survey
Number of items	18
Target population	Grade 6–12 students
Instrument source	Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of children's knowledge and regulation of cognition. *Contemporary Educational Psychology, 27,* 51–79.
Past administration of instrument: In school setting?	Yes
Past administration of instrument: In secondary school setting?	Yes
Past administration of instrument: Uses?	Formative
Reliability	Cronbach's α was .82 for the total scale, .76 for the Knowledge of Cognition subscale, and .80 for the Regulation of Cognition subscale. The conventionally accepted criterion of reliability for Cronbach's α is $\geq$ .70 (Nunnally, 1978).
Content validity	The authors outlined a theory that defines the measured skills.
Substantive validity	Not available
Structural validity	The authors conducted a confirmatory factor analysis, which showed that the measure had two underlying factors corresponding to Knowledge of Cognition and Regulation of Cognition. These findings are consistent with the results of previous studies (Kirbulut, 2014; Sperling et al., 2002). Fit indices for this model were within conventionally accepted criteria.
External validity	Not available
Generalizability	Not available
Consequential validity	Students' scores on the 18 item version of the measure correlated significantly with their scores on the Swanson Metacognitive Questionnaire (Swanson, 1990), their science grade point average, and their overall grade point average (Sperling et al., 2002).
Fairness	Not available

Source: Kim, B., Zyromski, B., Mariani, M., Lee, S., & Carey, J. (2017). Establishing the factor structure of the 18-item version of the Junior Metacognitive Awareness Inventory. *Measurement and Evaluation in Counseling and Development, 50*(1–2), 48–57.

Table B12. Self-Directed Learning Inventory: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Self-regulated learning
Format	Student self-report survey
Number of Items	10
Target population	Middle school, high school, and college students
Instrument source	Lounsbury, J., Levy, J., Park, S., Gibson, L., & Smith, R. (2009). An investigation of the construct validity of the personality trait of self-directed learning. *Learning and Individual Differences, 19*(4), 411–418.
Past administration of instrument: In school setting?	Yes
Past administration of instrument: In secondary school setting?	Yes
Past administration of instrument: Uses?	Research
Reliability	In the middle and high school samples the Cronbach's α was .87. The conventionally accepted criterion of reliability for Cronbach's α is $\geq$ .70 (Nunnally, 1978).
Content validity	The authors outlined a theory that defines the measured skills.
Substantive validity	Not available
Structural validity	The authors conducted a confirmatory factor analysis, which determined that a one-factor model adequately fit the 10 items.
External validity	The authors reported correlations between the Self-Directed Learning scale and measures for personality, satisfaction, interest, and aptitude. Measures reported were too extensive to list individually; the correlations can be found on p. 414 of the linked article.
Generalizability	Not available
Consequential validity	Correlations between Self-Directed Learning and cumulative grade point average were r = .26 for grade 9, r = .26 for grade 10, and r = .37 for grade 12.
Fairness	Not available

Source: Lounsbury, J., Levy, J., Park, S., Gibson, L., & Smith, R. (2009). An investigation of the construct validity of the personality trait of self-directed learning. *Learning and Individual Differences, 19*(4), 411–418.

Table B13. Self-Regulation Strategy Inventory—Self-Report: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Self-regulated learning
Format	Student self-report survey
Number of items	28
Target population	High school students
Instrument source	Cleary, T. (2006). The development and validation of the Self-Regulation Strategy Inventory—Self-Report. *Journal of School Psychology, 44*(4), 307–322.
Past administration of instrument: In school setting?	Yes
Past administration of instrument: In secondary school setting?	Yes
Past administration of instrument: Uses?	Formative
Reliability	Cronbach's α for the overall instrument was .92, with the subscales ranging from .72 to .88. The conventionally accepted criterion of reliability for Cronbach's α is $\geq$ .70 (Nunnally, 1978).
Content validity	The author outlined a theory that defines the measured skills.
Substantive validity	Not available
Structural validity	The principal component analysis yielded a three-factor structure that includes Seeking and Learning Information, Managing Environment/Behavior, and Maladaptive Regulatory Behaviors. Statistics indicating whether the data fit a pre-specified or theoretical factor structure were not provided.
External validity	Self-motivational beliefs positively predicted the self-regulation factors from the instrument.
Generalizability	Not available
Consequential validity	Univariate analyses showed a relationship between high- and low-achieving performance in science and high and low scores for each of the factors in the instrument.
Fairness	Not available

Source: Cleary, T. (2006). The development and validation of the Self-Regulation Strategy Inventory—Self-Report. *Journal of School Psychology, 44*(4), 307–322.

Perseverance and self-regulated learning

Three instruments were designed to measure perseverance and self-regulated learning (tables B14–B16).

Table B14. Motivated Strategies for Learning Questionnaire: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Perseverance and self-regulated learning
Format	Student self-report survey
Number of items	81
Target population	Developed for college students but administered to high school students
Instrument source	Pintrich, P., Smith, D., García, T., & McKeachie, W. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).* Ann Arbor, MI: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.
Past administration of instrument: In school setting?	Yes
Past administration of instrument: In secondary school setting?	Yes
Past administration of instrument: Uses?	Formative
Reliability	Cronbach's α values range from .52 to .93. The conventionally accepted criterion of reliability for Cronbach's α is $\geq$ .70 (Nunnally, 1978).
Content validity	The authors outlined a theory that defined the measured skills.
Substantive validity	Not available
Structural validity	The authors performed a confirmatory factor analysis in an attempt to show that the measure contained 15 unique factors: Intrinsic Goal Orientation, Control Beliefs About Learning, Extrinsic Goal Orientation, Self-Efficacy For Learning And Performance, Task Value, Test Anxiety, Rehearsal, Effort Management, Peer Learning, Elaboration, Metacognition, Organization, Critical Thinking, Help Seeking, and Time and Study Environments. Fit statistics for the factor model did not meet conventionally accepted criteria for good fitting models.
External validity	Factor correlations were reported from the confirmatory factor analysis model. Most correlations were in the expected direction.
Generalizability	Not available
Consequential validity	Not available
Fairness	Not available

Source: Pintrich, P., Smith, D., García, T., & McKeachie, W. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).* Ann Arbor, MI: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

Table B15. Program for International Student Assessment Student Learner Characteristics as Learners: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Perseverance and self-regulated learning
Format	Student self-report survey
Number of items	49
Target population	High school students
Instrument source	Artlet, C., Baumert, J., Julius-McElvany, N., & Peschar, J. (2003). *Learners for life: Student approaches to learning: Results from PISA 2000*. Paris, France: Organisation for Economic Co-operation and Development. https://eric.ed.gov/?id=ED480899.
Past administration of instrument: In school setting?	Yes
Past administration of instrument: In secondary school setting?	Yes
Past administration of instrument: Uses?	Research
Reliability	Cronbach's α values range from .74 to .86 for the U.S. sample. The conventionally accepted criterion of reliability for Cronbach's α is $\geq$ .70 (Nunnally, 1978).
Content validity	The authors outlined a theory that defines the measured skills.
Substantive validity	Not available
Structural validity	The authors hypothesized that the measure contains 13 factors: Memorization, Elaboration, Control, Instrumental Motivation, Interest in Reading, Interest in Mathematics, Effort and Persistence, Self-efficacy, Self-concept in Reading, Mathematical Self-concept, Academic Self-concept, Cooperative Learning, and Competitive Learning. The factor model exhibited fit statistics within conventionally accepted criteria.
External validity	Not available
Generalizability	Not available
Consequential validity	Not available
Fairness	Not available

Source: Artlet, C., Baumert, J., Julius-McElvany, N., & Peschar, J. (2003). *Learners for life: Student approaches to learning: Results from PISA 2000*. Paris, France: Organisation for Economic Co-operation and Development. https://eric.ed.gov/?id=ED480899.

Table B16. Student Engagement Instrument: Summary of reliability and validity information

Type of information	Summary
Social and emotional learning skill	Perseverance and self-regulated learning
Format	Student self-report survey
Number of items	56
Target population	High school students
Instrument source	Appleton, J., Christenson, L., Kim, D., & Reschly, A. (2006). Measuring cognitive and psychological engagement: Validation of the student engagement instrument. *Journal of School Psychology, 44*(5), 427–445.
Past administration of instrument: In school setting?	Yes
Past administration of instrument: In secondary school setting?	Yes
Past administration of instrument: Uses?	Research
Reliability	For the six factors, Cronbach's α values were .88 for Teacher–Student Relationships, .80 for Control and Relevance of School Work, .82 for Peer Support for Learning, .78 for Future Aspirations and Goals, .76 for Family Support for Learning, and .72 for Extrinsic Motivation. The conventionally accepted criterion of reliability for Cronbach's α is $\geq$ .70 (Nunnally, 1978).
Content validity	The authors outlined a theory that defines the measured skills.
Substantive validity	Not available
Structural validity	The authors first conducted an exploratory factor analysis with half of the data and then sought to confirm the factor structure out of sample with the other half of the data. The authors suggest the items measured six unique factors: Teacher–Student Relationships, Control and Relevance of School Work, Peer Support for Learning, Future Aspirations and Goals, Family Support for Learning, and Extrinsic Motivation. The confirmatory factor analysis model exhibited fit within conventionally accepted criteria.
External validity	The engagement factors were correlated with each other in the expected direction.
Generalizability	Not available
Consequential validity	Positive relationships were noted between most Student Engagement Instrument factors and academic indicators such as grade point average and reading and math scores, and negative relationships were noted between most of the Student Engagement Instrument factors and school suspension.
Fairness	Not available

Source: Appleton, J., Christenson, L., Kim, D., & Reschly, A. (2006). Measuring cognitive and psychological engagement: Validation of the student engagement instrument. *Journal of School Psychology, 44*(5), 427–445.

Notes

1. Fairness is listed as a component of validity based on recommendations from the Standards for Educational and Psychological Testing, the gold standard for understanding test validity, which states that "fairness is a fundamental validity issue and requires attention throughout all stages of test development and use" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014, p. 49).
2. Total counts provided are not mutually exclusive. For example, some of the 30 instruments that were excluded because they had not been administered with secondary school students in the United States were also not publicly available.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Appleton, J., Christenson, L., Kim, D., & Reschly, A. (2006). Measuring cognitive and psychological engagement: Validation of the student engagement instrument. *Journal of School Psychology, 44*(5), 427–445.

Artlet, C., Baumert, J., Julius-McElvany, N., & Peschar, J. (2003). *Learners for life: Student approaches to learning: Results from PISA 2000.* Paris, France: Organisation for Economic Co-operation and Development. https://eric.ed.gov/?id=ED480899.

Atkins-Burnett, S. A., Fernández, C., Akers, L., Jacobson, J., & Smither-Wulsin, C. (2012). *Landscape analysis of non-cognitive measures.* Princeton, NJ: Mathematica Policy Research. Retrieved July 19, 2017, from https://www.mathematica-mpr.com/our-publications-and-findings/publications/landscape-analysis-of-noncognitive-measures.

Champlain Valley School District. (2016). *Champlain Valley School District mission statement.* Hinesburg, VT: Author. Retrieved July 19, 2017, from https://drive.google.com/file/d/0B8p5Y8F0P258N1lma2hKSWtBd2M/view.

Cleary, T. (2006). The development and validation of the Self-Regulation Strategy Inventory—Self-Report. *Journal of School Psychology, 44*(4), 307–322.

Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy & Practice, 3*(3), 265–286.

Duckworth, A., Peterson, C., Matthews, M., & Kelly, D. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*(6), 1087–1101.

Duckworth, A., & Quinn, P. (2009). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment, 91*(2), 166–174.

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.

Durlak, J. A., Domitrovich, C. E., Weissberg, R. P., & Gullotta, T. P. (Eds.). (2015). *Handbook of social and emotional learning: Research and practice.* New York, NY: Guilford Publications.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*(1), 405–432.

Eccles, J., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 1017–1095). New York, NY: Wiley.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching adolescents to become learners—The role of noncognitive factors in shaping school performance: A critical literature review.* Chicago, IL: University of Chicago Consortium on School Research.

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 763–782). Boston, MA: Springer USA.

Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). *Measuring student engagement in upper elementary through high school: A description of 21 instruments.* (Issues & Answers Report, REL 2011–098). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. http://ies.ed.gov/ncee/edlabs.

French, B., Gotch, C., Immekus, J., & Beaver, J. (2016). An investigation of the psychometric of a measure of teamwork among high school students. *Psychological Test and Assessment Modeling, 58*(3), 455–470.

Gayl, C. L. (2017). *How state planning for the Every Student Succeeds Act (ESSA) can promote student academic, social, and emotional learning: An examination of five key strategies.* Chicago, IL: Collaborative for Academic Social, and Emotional Learning (CASEL). Retrieved July 19, 2017, from http://www.casel.org/wp-content/uploads/2017/04/ESSA-and-SEL-Five-Strategies-April-2017–041717.pdf.

Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice, 23*(1), 17–27.

Heckman, J. J. (2008). Schools, skills, and synapses. *Economic Inquiry, 46*(3), 289–324.

Howard, B., McGee, S. Shia, R., & Hong, N. (2000, April). *Metacognitive self-regulation and problem-solving: Expanding the theory base through factor analysis.* Paper presented at the Annual Meeting of the American Educational Research Association, April 24–28, 2000, New Orleans, LA. https://eric.ed.gov/?id=ED470973.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: A multidisciplinary journal, 6*(1), 1–55.

Johnson, R. T., & Johnson, D. W. (1994). An overview of cooperative learning. In J. S. Thousand, R. A. Villa, & A. I. Nevin (Eds.), *Creativity and collaborative learning: A practical guide to empowering students and teachers* (pp. 31–44). Baltimore, MD: Brookes Press.

Kafka, T. (2016). *A list of non-cognitive assessment instruments.* New York, NY: Community College Research Center, Teachers College, Columbia University. Retrieved July 19, 2017, from https://ccrc.tc.columbia.edu/images/a-list-of-non-cognitive-assessment-instruments.pdf.

Kim, B., Zyromski, B., Mariani, M., Lee, S., & Carey, J. (2017). Establishing the factor structure of the 18-item version of the Junior Metacognitive Awareness Inventory. *Measurement and Evaluation in Counseling and Development, 50*(1–2), 48–57.

Kirbulut, Z. D. (2014). Modeling the relationship between high school students' chemistry self-efficacy and metacognitive awareness. *International Journal of Environmental and Science Education, 9*(2), 177–196.

Kosovich, J., Hulleman, C., Barron, K., & Getty, S. (2014). A practical measure of student motivation: Establishing validity evidence for the Expectancy-Value-Cost Scale in middle school. *Journal of Early Adolescence, 27*(1), 49–64.

Kuhn, D. (2015). Thinking together and alone. *Educational Researcher, 44*(1), 46–53.

Lounsbury, J., Levy, J., Park, S., Gibson, L., & Smith, R. (2009). An investigation of the construct validity of the personality trait of self-directed learning. *Learning and Individual Differences, 19*(4), 411–418.

Lower, L., Newman, T., & Anderson-Butcher, D. (2015). Validity and reliability of the Teamwork Scale for Youth. *Research on Social Work Practice, 27*(6), 1–10. Retrieved July 19, 2018, from https://www.researchgate.net/publication/282806754_Validity_and_Reliability_of_the_Teamwork_Scale_for_Youth.

Marks, H. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal, 37*(1), 153–184.

Melnick, H., Cook-Harvey, C. M., & Darling-Hammond, L. (2017). Encouraging social and emotional learning in the context of new accountability. Stanford, CA: Learning Policy Institute.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749. Retrieved October 12, 2018, from https://www.d.umn.edu/~kgilbert/ened5560–1/Validity%20-%20Messick1995.pdf.

Muis, K. R., Winne, P. H., & Jamieson-Noel, D. (2007). Using a multitrait–multimethod analysis to examine conceptual similarities of three self- regulated inventories. *British Journal of Educational Psychology, 77*(1), 177–195. Retrieved January 25, 2018, from https://onlinelibrary.wiley.com/doi/abs/10.1348/000709905X90876.

National Center on Safe Supportive Learning Environments. (2017). *School Climate Survey Compendia.* Washington, DC: Author. Retrieved July 19, 2017, from

https://safesupportivelearning.ed.gov/topic-research/school-climate-measurement/school-climate-survey-compendium.

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

O'Conner, R., De Feyter, J., Carr, A., Luo, J. L., & Romm, H. (2017). *A review of the literature on social and emotional learning for students ages 3–8: Characteristics of effective social and emotional learning programs (part 1 of 4)* (REL 2017–245). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. http://ies.ed.gov/ncee/edlabs

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*(4), 667–686. Retrieved January 25, 2018, from https://ueeval.ucr.edu/teaching_practices_inventory/Pintrich_2003.pdf.

Pintrich, P., Smith, D., García, T., & McKeachie, W. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor, MI: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

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.

Seifert, T. L. (2004). Understanding student motivation. *Educational Research, 46*(2), 137–149. Retrieved January 25, 2018, from https://www.tandfonline.com/doi/abs/10.1080/0013188042000222421.

Shechtman, N., DeBarger, A. H., Dornsife, C., Rosier, S., & Yarnall, L. (2013). *Promoting grit, tenacity, and perseverance: Critical factors for success in the 21st century*. Washington, DC: US Department of Education, Department of Educational Technology. Retrieved January 25, 2018, from http://pgbovine.net/OET-Draft-Grit-Report-2-17-13.pdf.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107.

Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of children's knowledge and regulation of cognition. *Contemporary Educational Psychology, 27*(1), 51–79.

Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of educational psychology, 82*(2), 306.

Wang, L., MacCann, C., Zhuang, X., Liu, L., & Roberts, R. (2009). Assessing teamwork and collaboration in high school students: A multimethod approach. *Canadian Journal of School Psychology, 24*(2), 41–54.

West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F. O., & Gabrieli, J. D. E. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis, 38*(1), 148–170.

What Works Clearinghouse. (2007). *What Works Clearinghouse topic report: Character education.* Washington, DC: Institute of Education Sciences, U.S. Department of Education. http://eric.ed.gov/?id=ED497054.

Zhuang, X., MacCann, C., Wang, L., Liu, O. L., & Roberts, R. D. (2008, October). *Development and validity evidence supporting a team-work and collaboration assessment for high school students* (Research Report No. RR-08–50). Ewing, NJ: ETS. Retrieved July 10, 2018, from https://pdfs.semanticscholar.org/f83e/641f4875466adbed23b353494bc0f6.a9d250.pdf.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA: Academic Press. Retrieved July 19, 2018, from https://www.sciencedirect.com/science/article/pii/B9780121098902500317.

The Regional Educational Laboratory Program produces 7 types of reports

Making Connections
Studies of correlational relationships

Making an Impact
Studies of cause and effect

What's Happening
Descriptions of policies, programs, implementation status, or data trends

What's Known
Summaries of previous research

Stated Briefly
Summaries of research findings for specific audiences

Applied Research Methods
Research methods for educational settings

Tools
Help for planning, gathering, analyzing, or reporting data or research