JOURNAL OF LEARNING ANALYTICS

Volume 7(1), 5—18. http://dx.doi.org/ 10.18608/jla.2020.71.2

Measuring Mastery Behaviours at Scale: The Persistence, Effort, Resilience, and Challenge-Seeking (PERC) Task

Tenelle Porter[1], Diego Catalán Molina[2], Lisa Blackwell[3], Sylvia Roberts[4], Abigail Quirk[5], Angela L. Duckworth[6], Kali Trzesniewski[7]

Abstract

Mastery behaviours — seeking out challenging tasks and continuing to work on them despite difficulties — are integral to achievement but difficult to measure with precision. The current study reports on the development and validation of the computer-based persistence, effort, resilience, and challenge-seeking (PERC) task in two demographically diverse samples of adolescents (total *N* = 3,188). We present evidence for convergent validity with self-reported mastery behaviours and learning mindsets, discriminant validity from theoretically unrelated constructs, and incremental predictive validity for grade point average (GPA). Results suggest that the PERC task may be a useful tool for assessing mastery behaviours at scale.

Notes for Practice (research paper)

- Mastery behaviours such as persisting through difficulty, exerting effort, striving after failure, and seeking challenge can promote success in school.
- The persistence, effort, resilience, and challenge-seeking (PERC) task assesses these four mastery behaviours in a single computer activity that takes about 10 minutes to complete and does not depend on language ability or subject-specific knowledge.
- Students with higher PERC scores have more adaptive learning mindsets and higher self-reported mastery behaviours.
- PERC scores also predict high school grade point average (GPA) independent of learning mindsets and demographic factors.
- PERC may be a useful tool for assessing mastery behaviours at scale.

Keywords

Performance tasks, persistence, effort, resilience, growth mindset, achievement

Submitted: 06.09.2019 — **Accepted:** 25.01.2019 — **Published:** 03.04.2020

Corresponding author [1]Email: tjporter@ucdavis.edu Address: Hart Hall UC Davis, 301 Shields Ave, Davis, CA 95616, United States
[2]Address: Hart Hall UC Davis, 301 Shields Ave, Davis, CA 95616, United States
[3]Address: 139 Van Dale Rd, Woodstock, NY 12498, United States
[4]Address: 6 Tam O Shanter Drive, Purchase, NY 10577, United States
[5]Address 3675 Market Street, Philadelphia, PA 19104, United States
[6]Address 3675 Market Street, Philadelphia, PA 19104, United States
Corresponding author [7]Email: ktrz@ucdavis.edu Address: Hart Hall UC Davis, 301 Shields Ave, Davis, CA 95616, United States

1. Introduction

Struggles are inherent to most important pursuits, yet they can lead to discouragement. Mastery-oriented behaviours include "the seeking of challenging tasks and the maintenance of effective striving under failure" (Dweck & Leggett, 1988, p. 256). These behaviours promote problem-solving (Diener & Dweck, 1978, 1980; Mueller & Dweck, 1998), learning (Licht & Dweck, 1984), and higher performance (Blackwell, Trzesniewski, & Dweck, 2007; Yeager et al., 2016). A meta-analysis has found that individual differences in mastery behaviours predict higher achievement in a range of domains, including school (Burnette, O'Boyle, VanEpps, Pollack, & Finkel, 2013). In the current investigation, we developed and validated a novel computer-based measure of mastery behaviours for adolescents, modelled after the experimental paradigm in Mueller & Dweck (1998), called the persistence, effort, resilience, and challenge-seeking (PERC) task.

To date, research on mastery behaviours has relied almost exclusively on questionnaires (see Burnette et al., 2013, for a compilation). Key benefits of questionnaires include their test-retest reliability, predictive validity, and cost-effectiveness (Enkavi et al., 2019; Vazire & Mehl, 2008). However, like any assessment approach, questionnaires have limitations. In particular, questionnaires rely on subjective judgments and are therefore vulnerable to a variety of response biases, including retrospective, acquiescence, and group reference bias (Abrahams et al., 2019). Questionnaires are also easy to fake (Arendasy, Sommer, Herle, Schützhofer, & Inwanschitz, 2011; Marcus, 2006; Ziegler, Schmidt-Atzert, Bühner, & Krumm, 2007).

The crucial advantage of performance tasks is that they capture actual behaviour. Scores do not typically depend on subjective judgments, nor are they prone to response biases or easily faked (Abrahams et al., 2019; Arendasy et al., 2011; Ziegler et al., 2007). Of course, performance tasks have their own limitations. Tasks tend to have lower test-retest reliability than questionnaires (Enkavi et al., 2019) and may be less effective than questionnaires at tapping typical versus maximal performance (Duckworth & Yeager, 2015; Willerman, Turner, & Peterson, 1976). Nevertheless, performance tasks complement questionnaires as part of an array of assessments (Abrahams et al., 2019).

2. What Are Mastery Behaviours?

The term *mastery behaviours* draws from research that identifies differences in how children respond to challenge, effortful work, and failure. A series of studies by Diener & Dweck (1978, 1980) that are summarized by Dweck & Leggett (1988) found that some children become anxious, depressed, or bored when they struggle, leading to ineffective strategies and giving up. In contrast, other children of comparable ability maintain positive affect and demonstrate effective strategies despite struggle and failure. This response was deemed *mastery oriented* (Diener & Dweck, 1978, 1980; Dweck & Leggett, 1988).

Research has typically focused on four related yet distinct mastery behaviours in childhood and adolescence. Challenge-seeking refers to selecting a more difficult task instead of an easier one (e.g., Yeager et al., 2016). Effort is investment in learning (e.g., Galla et al., 2014). Persistence is sustained engagement with a difficult task (e.g., Ventura, Shute, & Zhao, 2013). Finally, resilience refers to the capacity to recover after failure (e.g., Mangels, Butterfield, Lamb, Good & Dweck, 2006).

3. Performance Tasks of Mastery Behaviours

Although we are not aware of validated performance tasks that assess mastery behaviours in aggregate, several tasks tapping their component features assess individual differences in adolescents and emerging adults. Here we describe a subset of such tasks for which there is published evidence of convergent and criterion validity.

Challenge-seeking is typically assessed by asking participants to choose between an easy and a difficult task. A computer-based measure called the Make-A-Math-Worksheet Task has been validated in a field study with adolescents (Yeager et al., 2016). Students select from problems that are “not very challenging and you probably will not learn very much,” “somewhat challenging and you might learn a medium amount,” and “very challenging but you might learn a lot” (Yeager et al., 2016, p. 380). The difference between the number of easy and difficult problems selected is the target measure (Yeager et al., 2016, although see Bettinger, Ludvigsen, Rege, Solli, & Yeager, 2018, for an alternative scoring of the same task).

Effort assessments differ from challenge-seeking tasks by measuring investment in learning. Two computer-based tasks have been validated in field studies with adolescents. The Academic Diligence Task uses distractions to influence learning engagement by pitting simple but skill-developing math problems against entertaining diversions (video clips and the video game *Tetris*) (Galla et al., 2014). The number of problems solved and time spent on the problems are the target measures. In contrast, the Posterlet Game gives feedback to shape users’ experience (Cutumisu, Blair, Chin, & Schwartz, 2015). Here, participants design posters using an online platform with the option to request feedback and revise. The number of times corrective feedback is requested and the number of revisions completed are the target measures.

Persistence tasks differ from effort and challenge-seeking ones by assessing duration of engagement on very difficult problems (e.g., accuracy rate around 20 to 30%). Several unique computer-based measures have been developed for adolescents and emerging adults. These include the Anagrams-and-Riddles Task (ART), which has an option to skip difficult questions (Ventura et al., 2013), and the Mirror Tracing Frustration Task (Meindl et al., 2019), which provides entertaining games and videos as an alternative to a nearly impossible tracing activity. In addition, Newton’s Playground teaches properties such as balance, mass, and gravity by having students build machines (e.g., pendulums and springboards) to solve problems (Ventura & Shute, 2013). In each of these tasks, time spent on the difficult problems provides a measure of persistence.

Finally, resilience tasks capture the capacity to recover following failure, operationalized as how well you do (e.g., number correct) after failure. We were unable to find validated tasks assessing resilience, as defined here. However, some researchers have repurposed the flanker task — originally designed to tap response inhibitions — to measure resilience (Moser, Schroder, Heeter, Moran, & Lee, 2011). Participants’ objective in this task is to click target stimuli while distractor stimuli compete for their attention. Feedback on correctness is provided, and level of accuracy following mistakes gauges individuals’ resilience.

Others have used number correct on difficult tests to measure resilience. Mangels and colleagues (2006) tested college students on their general knowledge of literature, geography, math, history, and other subjects. After finishing the test, students saw which problems they answered right or wrong (the average score was around 40% correct) and the correct answers to each question. At the end of the session, researchers administered a surprise test on the missed questions. Performance on this test provided a measure of resilience.

4. Development of the PERC Task

Many researchers and educators are interested in not just one but a set of mastery behaviours. Accordingly, several self-report questionnaires assess mastery behaviours in aggregate (e.g., Blackwell et al., 2007; Elliot, McGregor, & Gable, 1999; Miller, Greene, Montalvo, Ravindran, & Nichols, 1996; Robins & Pals, 2002). Likewise, our goal was to capture an array of mastery responses in a single task. Inspired by Mueller & Dweck (1998), we adapted their experimental paradigm to create a measure of individual differences in mastery behaviours for adolescents. In so doing, we follow a tradition of adapting tasks originally designed for experiments to instead assess individual differences (see Knafo et al., 2007; Mischel, Shoda, & Peake, 1988; Ventura et al., 2013).

In Mueller & Dweck (1998), children completed three sets of Raven's matrices, which are nonverbal puzzles that measure abstract reasoning. In the first set, they did 10 medium-difficulty puzzles (an average of 52% correct). Next, they were given encouraging feedback, after which challenge-seeking was assessed by choosing from "problems that aren't too hard, so I don't get many wrong" and three similar options, and "problems that I'll learn a lot from, even if I won't look so smart" (Mueller & Dweck, 1998, p. 35). Four minutes was then allocated to solving the second set of more difficult puzzles (an average of 16% correct), after which persistence was assessed by asking "how much would you like to take these problems home to work on?" (Mueller & Dweck, 1998, p. 36). Finally, a third set of puzzles matched for difficulty with the first set was presented, the score on which was a measure of resilience.

For the PERC task, we administered four sets of Raven's matrices. We chose easy puzzles (80 to 90% correct) for the first set to create an experience of success that did not depend on predetermined feedback. We assessed challenge-seeking after the initial set of puzzles as a choice between easier and more difficult puzzles. The second set contained medium-difficulty puzzles (40 to 50% correct) with feedback and a chance to view tips that allowed us to assess effort. The third set of puzzles contained difficult puzzles (15 to 25% correct), allowing us to assess persistence as duration of engagement with difficult problems, and the final set was matched for difficulty with the first set, allowing us to assess resilience as post-failure performance.

Unlike Mueller & Dweck (1998), we were able to put PERC on the computer to make it self-directed. PERC is similar to learning technologies because it extracts trace data from users for analysis. It can also be considered a learning analytic tool because it facilitates "measurement, collection, analyses, and reporting of data about learners and their contexts for the purposes of understanding and optimizing learning" (Siemens, 2013, p. 1382).

However, PERC differs from some learning technologies in two key ways. First, PERC is a standalone assessment. This contrasts with learning technologies that measure socioemotional competencies as students advance through lessons in a system, class, or curriculum (e.g., Krumm et al., 2016). Because it is standalone, PERC can be used in a variety of classes to compare effects of different interventions or curricula.

Second, we designed PERC with *a priori* measurement goals in mind. This differs from data-mining, analysis of content and trace data from interactions with learning systems, and analysis of metadata from subject-specific assessments, which capture information about socioemotional skills from instruments that were not originally designed to measure these skills (e.g., Gibson, Kitto, & Bruza, 2016; Martin, Nacu, & Pinkard, 2016; Soland, Zamarro, Cheng, & Hitt, 2019).

5. Current Investigation

We conducted two studies to assess the validity of the PERC task. Study 1 was a smaller cross-sectional study in which we assessed convergence with self-reported *mastery behaviours* (i.e., exertion of effort, persistence, resilience, and challenge-seeking in school) and *growth mindset*, which is the belief that intelligence can be increased. Study 2 used a prospective design and a larger sample of adolescents to assess convergent validity with growth mindset; *failure-can-be-enhancing mindset*, which is the belief that failures can be beneficial for learning; and *academic intrinsic motivation*, which is interest in learning. We assessed convergence with these learning mindsets and attitudes because of research linking them to higher displays of mastery behaviours (Bettinger et al., 2018; Blackwell et al., 2007; Dweck & Leggett, 1988; Dweck, Chiu, & Hong, 1995; Dweck, 2000; Haimovitz & Dweck, 2016; Mueller & Dweck, 1998; Yeager et al., 2016). In addition, we assessed discriminant validity from theoretically unrelated constructs: closeness with friends, social belonging, and life satisfaction. Finally, we tested the incremental predictive validity of PERC for grade point average (GPA) over and above demographics, baseline puzzle-solving ability, and self-report questionnaires. Data and script are available for download at https://osf.io/8xcjp/.

6. Study 1

Our first aim was to assess the convergent validity of the PERC task with self-reported mastery behaviours and growth mindset.

6.1. Method

6.1.1. Participants and Procedure

In March 2017, we recruited 210 students in grades 6 to 8 (*M* age = 13.49, *SD* = .50) through Qualtrics Panels, a U.S. national survey company that recruits participants for online studies. Participants were ethnically diverse: 53% White, 17% Black, 10% Hispanic, 9% Asian, and 11% other ethnicity; 56% were female. Students completed the PERC task and self-report questionnaires in a single session.

6.1.2. Measures

PERC Task. The first screen in this computer-based task introduced the activity: "Strong puzzle solving skills are useful for many things. They can help you see problems in new ways and find creative ways of solving them too. Find out how strong your current puzzle solving skills are now."

Next, students were presented with four easy puzzles (Figure 1), whose percent correct score was used to assess *baseline puzzle-solving ability* (a similar measure was used in Mueller & Dweck, 1998).

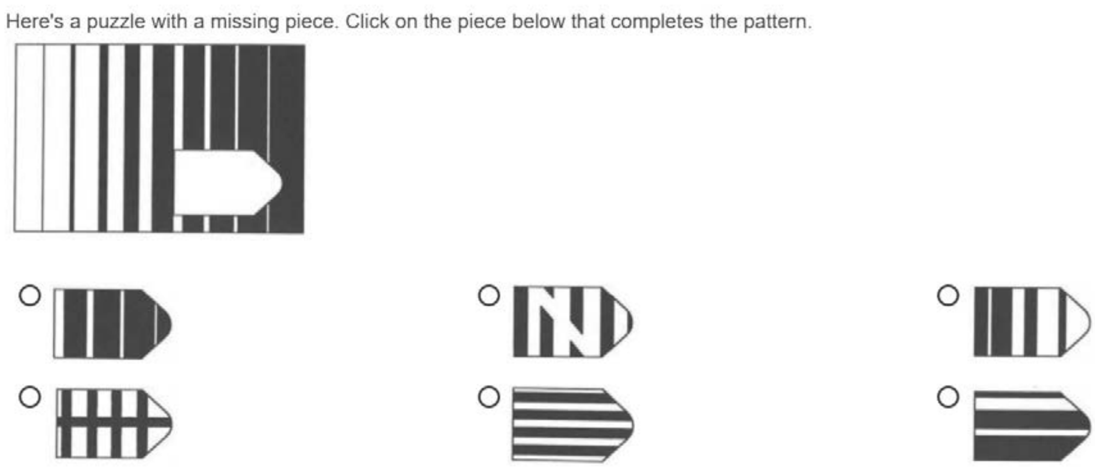

Figure 1. **Screenshot of baseline puzzle-solving ability section.**

The following screen read, "For the next set of puzzles, which kind would you like?" *Challenge-seeking* was coded zero if students chose "puzzles at the same level as the first set (kind of easy)" and one if they chose "puzzles that are more challenging (harder)."

The next screen read "Okay. Let's do a mix of harder and easier puzzles for now. On this set, we will tell you whether you get the answer right or wrong. Then, if you want, you will have a chance to read some tips for solving the ones you missed, and the ones you got right." Students were then presented with three medium-difficulty puzzles with the option to view puzzle-solving tips (Figure 2). *Effort* was measured as the total time spent on the puzzles and tips in this section, capturing both effort invested to solve the medium-difficulty problems and effort invested in reviewing the problem-solving tips[1].

[1] We also considered the possibility that fast and accurate completion might be spuriously labelled low effort, but in fact the correlation between time spent on the puzzles and score was positive: Pearson $r = .31$, $p < .001$. Alternative scorings of effort produced the same result (Supplementary Online Materials, Section 7).

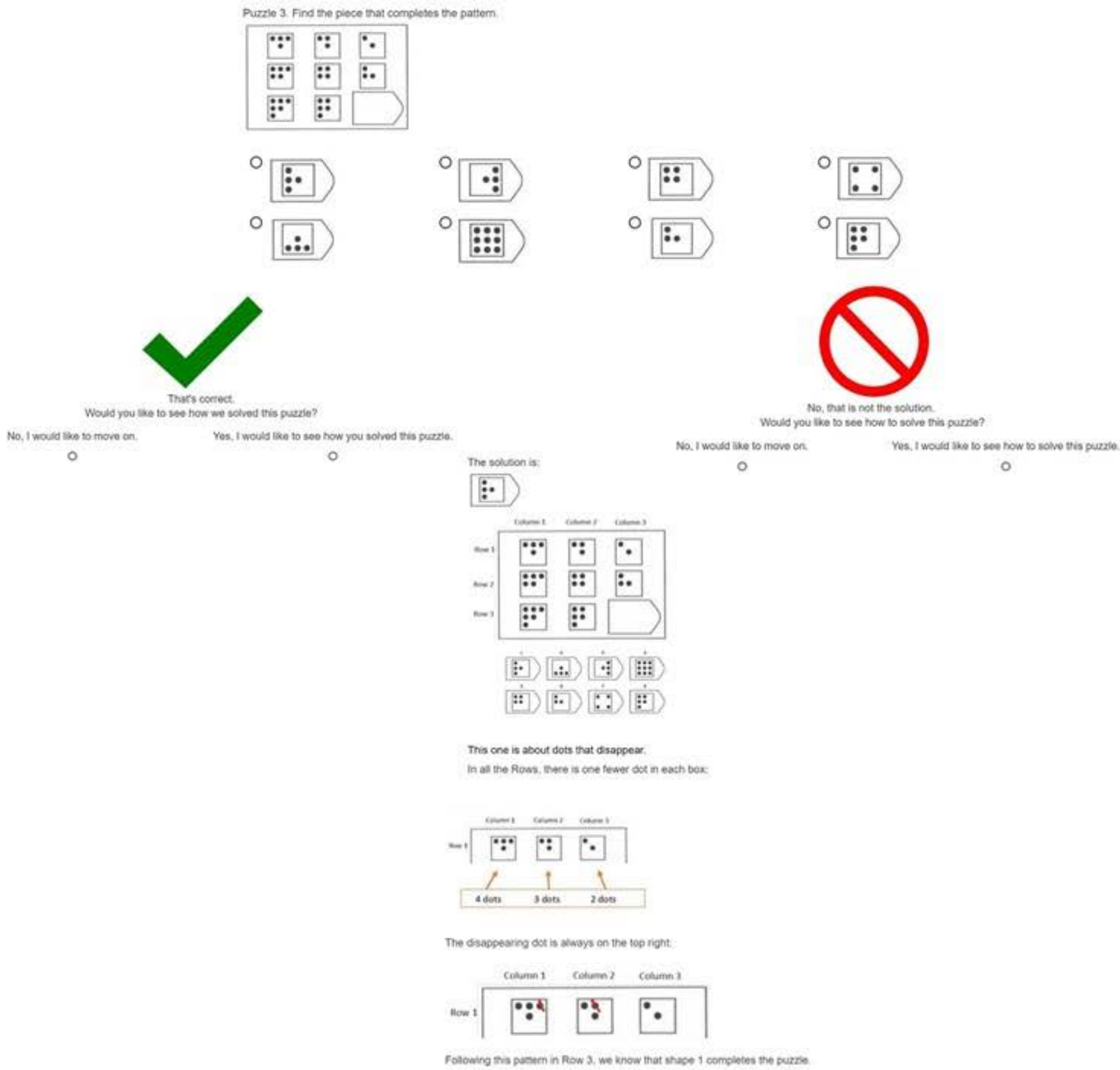

***Figure 2*. Screenshots of effort section.**

The next screen had an encouraging message displayed with a puppy: "You're about to start the final set of Puzzlemania! Here's a puppy to make you smile." The next screen read, "Welcome to the final set. It has 8 questions. You may do as many questions as you choose. If you choose to quit this set, you will move on to the next task. You will not be able to go back to the puzzles. Click next to begin." For a break from the prior set of medium-difficulty puzzles, students were then presented with one easy puzzle. Next, students were invited to work on four difficult puzzles (Figure 3), the total time on which was used to assess *persistence*. Finally, students were invited to work on three easy puzzles, the percent correct on which was a measure of *resilience*.

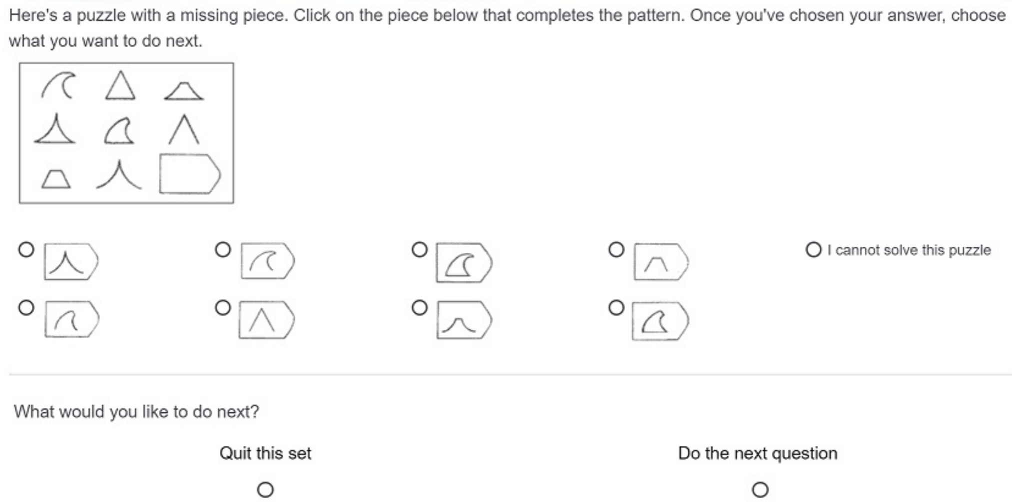

Figure 3. **Screenshot of persistence section.**

Puzzles were selected on the basis of difficulty. Difficulty was judged using puzzle rankings provided in the Raven's manual (Raven, Raven, & Court, 1998), and by pilot-testing (for more detail, see the Supplementary Online Materials, SOM, Section 2). Several alternative ways of scoring effort, persistence, and resilience produced the same results (SOM, Section 7). (To view the PERC task, see https://ucdavis.co1.qualtrics.com/jfe/form/SV_1R0rvSYoHMoH4mV.)

Self-Report Questionnaires. *Mastery behaviours* were measured with seven items adapted from Robins & Pals (2002), rated from 1 = *strongly disagree* to 5 = *strongly agree*: "I put a lot of effort into my school work"; "When I work on school assignments I put in a lot of effort"; "I do not give up when school work is difficult"; "When I am doing a challenging problem in school, I do not give up until I have found a solution"; "I choose to do more difficult activities in school instead of easy ones"; "I prefer doing challenging problems instead of ones that are easy to solve"; "If I get a disappointing grade on an assignment, I do my best to improve on the next one." Exploratory factor analysis supported a unidimensional structure (all item loadings ≥ .40) (Clark & Watson, 2019). Items had an internal reliability coefficient of $\alpha = .86$.

Growth mindset was measured with four items adapted from Blackwell et al. (2007), rated from 1 = *strongly agree* to 6 = *strongly disagree*: "You cannot change how smart you are"; "How smart you are is something about you that you can't change very much"; "People are born smart or not smart. This can't be changed"; and "You can learn new things, but you cannot make yourself smarter." Items had an internal reliability coefficient of $\alpha = .76$.

6.2. Results

The PERC puzzles were appropriately difficult. Students correctly solved an average of 89.52% ($SD = .20$) of the first set of easy puzzles, 37.62% ($SD = .26$) of the medium-difficulty puzzles, and 28.45% ($SD = .25$) of the difficult puzzles. These and remaining analyses were done with SPSS 25.

Because the effort and persistence subscores were positively skewed, we used a square root transformation and Winsorized outlier values to three standard deviations from the transformed mean. We then created a more parsimonious total *PERC score* for each student. Informed by Cohen, Cohen, Aiken, & West (1999), we rescaled the transformed effort and persistence subscores to a zero to one range[2]. To calculate the total PERC scores, we added to these effort and persistence subscores the binary challenge-seeking subscore and the percent correct resilience subscore. PERC scores therefore had a maximum possible value of four.

We report Pearson correlation coefficients for continuous variables and point-biserial correlations for dichotomous variables. As seen in Table 1, PERC subscores were intercorrelated (average $r = .34$, $p < .001$). Correlations for the subscores were, in general, directionally consistent but smaller in magnitude than the PERC total score. We therefore proceeded with the PERC score, controlling for age, gender, ethnicity, and baseline puzzle-solving ability in subsequent analyses testing convergent validity.

[2] The rescaling calculation took time for each student minus the minimum time, divided by the maximum minus the minimum time. The rescaled variables correlated at one with the originals, and the intervals were held constant between the original and rescaled measures.

Table 1: Descriptive Statistics and Intercorrelations of Main Study 1 Variables

	1	2	3	4	5	6	7	8
1. Total PERC score		$.71^{**}$	$.50^{**}$	$.70^{**}$	$.71^{**}$		$.28^{**}$	$.19^{**}$
2. Challenge-seeking subscore	$.73^{**}$		$.26^{**}$	$.19^{**}$	.08		$.18^{*}$	$.14^{*}$
3. Effort subscore	$.57^{**}$	$.32^{**}$		$.44^{**}$	$.14^{*}$		$.23^{**}$	.11
4. Persistence subscore	$.73^{**}$	$.25^{**}$	$.50^{**}$		$.61^{**}$		$.30^{**}$	$.18^{*}$
5. Resilience subscore	$.73^{**}$	$.14^{*}$	$.23^{**}$	$.63^{**}$			$.14^{*}$	.10
6. Baseline puzzle-solving ability	$.38^{**}$	$.29^{**}$	$.25^{**}$	$.23^{**}$	$.25^{**}$			$.26^{**}$
7. Self-reported mastery behaviours	$.33^{**}$	$.22^{**}$	$.30^{**}$	$.34^{**}$	$.18^{**}$	.13		
8. Growth mindset	$.27^{**}$	$.20^{**}$	$.18^{*}$	$.23^{**}$	$.16^{*}$	$.20^{**}$	$.29^{**}$	
Demographics								
Female	$.16^{*}$	.01	$.15^{*}$	$.19^{**}$	$.18^{**}$	.01	.08	.06
Age	.03	.06	.04	−.05	.00	.03	.02	−.01
White	.13	.05	.08	.01	$.17^{*}$	$.18^{**}$	.02	.12
Black	−.02	.02	−.01	.05	−.09	−.07	−.03	−.04
Hispanic	.04	.03	.10	.08	−.02	$-.16^{*}$	$.15^{*}$	.00
Asian	−.07	.00	−.06	−.06	−.10	.10	.03	−.04
Other ethnicity	−.12	−.13	−.11	−.08	−.03	−.08	−.11	−.10
M	1.87	0.55	0.39	0.30	0.63	0.90	3.79	4.02
SD	0.91	0.5	0.17	0.17	0.45	0.20	0.89	0.94
Observed Range	0.05–3.53	0, 1	0.00–1.00	0.00–1.00	0.00–1.00	0.00–1.00	1.00–5.00	1.00–5.00

Note. Partial correlations controlling for age, gender, ethnicity, and baseline puzzle-solving ability are presented above the diagonal, and correlations not controlling for covariates are presented below the diagonal. $^{*}p < .05$; $^{**}p < .01$.

PERC scores were positively correlated with both self-report mastery behaviours (partial correlation, $pr = .28$, $p < .001$), and growth mindset ($pr = .19$, $p < .001$). Findings provide initial evidence for the convergent validity of PERC as a measure of mastery behaviours.

7. Study 2

Study 2 was a prospective examination of PERC's properties in a larger sample of high school students. We tested convergence with growth mindset, beliefs about failure and academic intrinsic motivation, discriminant validity with theoretically unrelated constructs, and the incremental predictive validity of PERC for report card grades over and above demographics and self-report measures.

7.1. Method

7.1.1. Participants and Procedure

A total of $N = 2,978$ students in grades 9 through 12 (*M* age = 16.42, *SD* = 1.15) from seven public high schools were recruited through the Character Lab Research Network (see SOM, Section 3, for more information on recruitment). Students were socioeconomically and ethnically diverse: 68% eligible for free or reduced price lunch, 42% Hispanic, 32% White, 18% Black, 5% Asian, and 3% other ethnicity; 53% were female.

In January 2018, a teacher proctor at each school introduced the study by reading a brief script (SOM, Section 4), after which students completed the 10-minute PERC task on school computers. In May 2018, the same protocol was used to administer a battery of self-report questionnaires. Brevity of protocol and short measures were used to minimize respondent burden and increase efficiency at scale (Yeager & Bryk, 2015).

7.1.2. Measures

The PERC task was the exact same as that administered in Study 1.

Growth mindset was measured with three items from Blackwell et al. (2007), rated from 1 = *strongly agree* to 6 = *strongly disagree*: "Your intelligence is something about you that you can't change very much"; "You have a certain amount of intelligence, and you really can't do much to change it"; and "You can learn new things, but you can't really change your basic intelligence." Items had an internal reliability coefficient of $\alpha = .88$.

Failure-can-be-enhancing mindset was measured with two items from Haimovitz & Dweck (2016), rated from 1 = *strongly agree* to 5 = *strongly disagree*: "The effects of failure are bad and should be avoided" and "I believe experiencing failure hurts my learning and growth." The Spearman-Brown coefficient was .68 (Eisinga, te Grotenhuis, & Pelzer, 2013). A third item, "I believe experiencing failures helps me learn and grow," reduced the internal consistency of the scale and was removed; results were not different when this item was included.

Academic intrinsic motivation was assessed with two items adapted from Gottfried (1985), rated from 1 = *strongly disagree* to 5 = *strongly agree*: "I like learning" and "Learning is personally important to me." The Spearman-Brown coefficient was .75.

Social belonging was measured with two items adapted from Walton & Cohen (2007), rated from 1 = *not at all* to 5 = *a great deal*: "How much do you feel like you fit in at school?" and "How much do you feel like you belong at school?" The Spearman-Brown coefficient was .83.

Life satisfaction was measured with a single item, "Overall, how do you feel about your life? Indicate how satisfied or dissatisfied you are these days," rated by moving a horizontal slider that made a smiley face graphic move progressively from frowning to smiling as it went from one to five. Single-item life satisfaction assessments are common in large studies (Diener, Inglehart, & Tay, 2013; Lucas & Donnellan, 2012).

Closeness with friends was assessed with an item adapted from Aron, Aron, & Smollan (1992) that asked, "Which picture best describes your relationship with a good friend?" Students chose among five Venn diagrams with circles labelled "friend" and "self" that progressively increased in overlap.

Finally, we obtained report card grades in all subjects from school records and averaged them to create a measure of end-of-year GPA for each student.

7.2. Results

The PERC puzzles were appropriately difficult for this sample of high school students. Students correctly solved an average of 86.68% (*SD* = .22) of the first set of easy puzzles, 40.32% (*SD* = .28) of the medium-difficulty puzzles, and 25.22% (*SD* = .26) of the difficult puzzles.

Replicating the findings from Study 1, PERC subscores were intercorrelated (average $r = .35$, $p < .001$) and their associations with other constructs were directionally consistent but smaller in magnitude than the PERC total score (Table 2). We therefore proceeded with the PERC score, controlling for age, gender, socioeconomic status, grade level, ethnicity, school, and baseline puzzle-solving ability, in subsequent analyses examining convergent validity.

Table 2: Descriptive Statistics and Intercorrelations of Main Study 2 Variables

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Total PERC score		.63**	.50**	.72**	.80**		.12**	.13**	.12**	.04	.00	−.01	.15**
2. Challenge-seeking subscore	.67**		.20**	.18**	.12**		.09**	.09**	.08**	.01	−.003	.003	.10
3. Effort subscore	.55**	.27**		.47**	.19**		.09**	.08**	.11**	.04*	.02	.01	.12
4. Persistence subscore	.74**	.25**	.50**		.59**		.09**	.10**	.11**	.04*	−.02	.01	.10
5. Resilience subscore	.79**	.18**	.23**	.61**			.07**	.08**	.07**	.03	.002	−.02	.10
6. Baseline puzzle-solving ability	.34**	.30**	.26**	.24**	.19**								
7. Growth mindset	.14**	.11**	.10**	.10**	.09**	.07**		.11**	.23**	.14**	.12**	.02	.12**
8. Failure-can-be-enhancing mindset	.16**	.12**	.11**	.13**	.10**	.13**	.12**		.05*	−.02	.06**	−.02	.09**
9. Academic intrinsic motivation	.10**	.07**	.10**	.10**	.06**	−.03	.22**	.04*		.32**	.22**	.04*	.25**
10. Social belonging	.02	.00	.03	.03	.02	−.08**	.13**	−.02	.31**		.43**	.18**	.15**
11. Life satisfaction	−.02	−.02	.01	−.03	.00	−.08**	.12**	.05**	.23**	.44**		.13**	.20**
12. Closeness with friends	.01	.02	.01	.02	−.01	.04*	.02	−.01	.03	.18**	.12**		−.02
13. GPA	.17**	.13**	.12**	.12**	.11**	.13**	.12**	.11**	.24**	.11**	.15**	.00	
Demographics													
Female	−.06**	−.06**	−.04*	−.04*	−.04*	.00	−.02	−.04*	.04*	−.12**	−.10**	.03	.16**
Free and reduced-price lunch	−.07**	−.10**	.03	−.05*	−.04*	−.12**	−.01	−.08**	.02	−.01	−.02	−.06**	−.18**
Age	−.01	−.03	.01	.00	.01	−.01	−.03	.01	.09**	.03	.06**	−.06**	−.03
Hispanic	−.07**	−.06**	.00	−.08*	−.04*	−.07**	−.03	−.08**	−.01	.01	.02	−.05*	−.11**
Black	−.03	−.07**	.04	.02	−.01	−.13**	.02	−.01	.03	.04*	.03	−.02	−.08**
White	.08**	.10**	−.04*	.05*	.05*	.15**	.02	.07**	−.03	−.02	−.01	.08**	.11**
Asian	.02	.01	.03	.03	.01	.03	.00	.03	.02	−.03	−.05*	−.04*	.12**
Other ethnicity	.01	.03	.00	.00	−.01	.01	.00	.01	.01	−.01	−.03	.00	.03
M	1.98	0.77	0.42	0.26	0.53	3.47	4.03	4.14	3.69	3.20	3.61	5.03	83.91
SD	0.88	0.42	0.17	0.17	0.48	0.86	1.24	1.00	0.85	1.02	1.20	1.83	8.39
Observed range	0.02–3.92	0, 1	0.00–1.00	0.00–1.00	0.00–1.00	0.00–4.00	1.00–6.00	2.00–6.00	1.00–5.00	1.00–5.00	1.00–5.00	1.00–7.00	54.60–100.00

Note. Partial correlations controlling for gender, free and reduced-price lunch status, age, ethnicity, school, and baseline puzzle-solving ability are presented above the diagonal, and correlations not controlling for covariates are presented below the diagonal. $^{*}p < .05$; $^{**}p < .01$.

As expected, PERC scores were positively correlated with growth mindset (partial correlation ($pr = .12$, $p < .001$)), failure-can-be-enhancing mindset ($pr = .13$, $p < .001$), and academic intrinsic motivation ($pr = .12$, $p < .001$). Likewise, PERC scores were not correlated with social belonging ($pr = .04$, $p = .05$), life satisfaction ($pr = .00$, *ns*), or closeness with friends ($pr = .00$, *ns*).

To assess the incremental predictive validity of PERC for final GPA, we fit an ordinary least squares regression model that controlled for self-report questionnaires in addition to demographics and school. As shown in Table 3, PERC scores predicted final GPA ($\beta = .12$, $p < .001$) in this model.

Table 3: Regression Analysis of Main Study Variables Predicting GPA

Variables	β	95% CI
Female	$.18^{**}$	[0.14, 0.21]
F/R lunch	$-.18^{**}$	[−0.23, −0.14]
Age	−.03	[−0.06, 0.01]
Ethnicity		
Hispanic	$-.09^{**}$	[−0.13, −0.04]
Black	$-.06^{**}$	[−0.11, −0.02]
Asian	$.08^{**}$	[0.04, 0.11]
Other	.01	[−0.03, 0.04]
School		
One	$-.19^{**}$	[−0.23, −0.14]
Two	.02	[−0.01, 0.06]
Three	$-.07^{**}$	[−0.12, −0.03]
Four	$-.06^{**}$	[−0.10, −0.02]
Five	$-.08^{**}$	[−0.12, −0.05]
Six	$-.06^{*}$	[−0.11, −0.02]
Baseline puzzle-solving ability	$.05^{**}$	[0.02, 0.09]
Self-report questionnaires		
Growth mindset	.03	[0.00, 0.07]
Failure-can-be-enhancing mindset	$.05^{**}$	[0.01, 0.08]
Academic intrinsic motivation	$.18^{**}$	[0.14, 0.22]
Social belonging	.03	[−0.01, 0.07]
Life satisfaction	$.13^{**}$	[0.10, 0.17]
Closeness with friends	$-.04^{*}$	[−0.08, −0.01]
Total PERC score	$\mathbf{.12^{**}}$	**[0.08, 0.15]**

Note. Adjusted $R^2 = .20$ ($n = 2{,}786$). CI = confidence interval for β. F/R lunch = Free or reduced-price lunch status. Ethnicities were dummy variables, with White as the reference category. Schools were dummy variables, with school 7 ($n = 769$) as the reference category.
$^{*}p < .05$; $^{**}p < .01$.

8. General Discussion

The current study introduces the PERC task as a novel performance measure of mastery behaviours. In PERC, students work on several sets of puzzles that are embedded in a coherent task narrative. They choose between puzzles that are easy and difficult (challenge-seeking), decide how long to practise (effort), and choose how long to spend on difficult puzzles (persistence), and they solve easier puzzles after struggling with difficult ones (resilience). Mastery behaviours are summed to create an aggregate PERC score for each individual.

We found that PERC puzzles were at the expected difficulty levels for two diverse samples of adolescents. PERC scores demonstrated convergent validity, with individual differences in self-report mastery behaviours, growth mindset, failure-can-be-enhancing mindset, and academic intrinsic motivation. PERC scores also showed evidence of discriminant validity: they were not correlated with social belonging, life satisfaction, or closeness with friends. Moreover, PERC had incremental predictive validity for objectively measured GPA, accounting for additional variance beyond that accounted for by demographics and self-report questionnaires.

The magnitude of effects was comparable to past findings. Intercorrelations with self-report questionnaires were small to medium in magnitude (pr from .12 to .28), which compares favourably to meta-analytic estimates of the association between questionnaire and behavioural measures of self-control (r = .10 to .21; Duckworth & Kern, 2011) and published correlations between the Academic Diligence Task and conscientiousness, grit, and self-control (pr = .08, .16, and .11; Galla et al., 2014) and the Make-A-Math-Worksheet Task and growth mindset (r = .13; Yeager et al., 2016). The incremental association between PERC scores and GPA (β = .12) was comparable to associations between GPA and the Academic Diligence task (β = .07 and .17; Galla et al., 2014) and the Mirror Tracing Frustration Task (β = .10 to .15; Meindl et al., 2019).

Funder & Ozer (2019) recommend gauging practical importance by comparing effect sizes with those that are commonly experienced. For example, the effect of an antihistamine on a runny nose and sneezing is around r = .11, and the effect of ibuprofen on pain is about r = .14 (Funder & Ozer 2019). In an educational context, the effect of receiving free breakfast on academic achievement is about r = .05, and the effect of having a personal tutor on achievement is about r = .11 (Kraft, 2019). The association between PERC scores and GPA is meaningful by comparison. Small behavioural effects can also accumulate over time, for example, through cumulative and interactional continuity (Caspi, Elder, & Bem, 1988; Funder & Ozer, 2019).

To our knowledge, PERC is the only performance task of aggregate mastery behaviours to demonstrate evidence of convergent, discriminant, and incremental predictive validity. As such, it has the potential to advance research on adolescents' mastery behaviours in several ways. PERC makes assessment of actual mastery behaviours feasible by requiring few resources to facilitate and little time to administer (i.e., median completion time was around seven minutes). It is therefore well suited for school-based studies for which resources and time are limited and thus could be administered multiple times to track development without unduly burdening schools and students. The PERC task may also be appropriate for populations that are diverse in terms of language, numeracy, and culture. The Raven's matrices used throughout the task do not depend on prior knowledge of math, science, language, or other subjects and have been used successfully in global contexts. (See Choudhury & Gorman (1999) for a study with Guatemalan adolescents and Costenbader & Ngari (2001) for a study with Kenyan children.)

9. Limitations and Future Directions

Like all studies, the current research has limitations. Although we tested PERC in two demographically diverse samples of American adolescents, additional studies are needed to examine its external validity (Henrich, Heine, & Norenzayan, 2010). These should include adolescents of different ages and living in other cultural contexts.

Second, we did not examine the test-retest reliability of PERC and therefore cannot report on its stability. In general, performance tasks tend to have lower test-retest reliability than questionnaires, in part because increased familiarity with the task and differences in mood and assessment environments can influence performance (Enkavi et al., 2019; Hedge, Powell, & Sumner, 2018). However, Hedge and colleagues (2018) note that lower test-retest reliability does not mean that tasks "are not *replicable*, *valid*, *or robust* measures of their respective constructs" (p. 1178, italics original). Rather, the inherent limitations and strengths of tasks should be considered in deciding when and how to use them (Duckworth & Yeager, 2015). For example, tasks may be especially sensitive to fluctuations in behaviour over time and in response to interventions (Enkavi et al., 2019).

Third, PERC has limitations relative to other learning technologies. The task is brief and does not provide as much information about students as longer games or system interactions provide. PERC is also administered outside of typical learning activities. This contrasts with learning management systems that assess behaviours without disrupting learning (e.g., Joksimović, Gašević, Loughin, Kovanović, & Hatala, 2015). The latter may capture higher-quality data about the learning process (Siemens, 2013), whereas PERC is designed to capture motivational behaviour in a controlled environment to examine individual differences across students (e.g., in response to an intervention). In addition, PERC measures time-on-task, which is not a perfect metric of engagement in learning (Kovanović et al., 2015), though we found that alternative scorings (e.g., counts instead of time) produced comparable results (see SOM, Section 7, for detailed findings). PERC is one tool at learning analysts' disposal — its strengths and limitations should be considered in relation to other learning technologies.

10. Conclusion

Adolescents are certain to encounter difficulties as they advance in school. In order to achieve to their highest potential, they will need to embrace challenge and continue to strive despite setbacks. The scientific study of mastery behaviours has, thus far, relied almost exclusively on questionnaires. As a complementary measurement approach, we developed a computer-based assessment of persistence, effort, resilience, and challenge-seeking (PERC). In the current study, PERC made it possible to collect behavioural data on a few thousand students in less time than a class period and, since no trained researchers were needed for administration, at near zero marginal cost. Individual differences in PERC scores converged with self-reported mastery behaviours, learning mindsets, and predicted GPA beyond self-report measures in a diverse sample of adolescents. We believe PERC can contribute to the comprehensive and multimethod study of mastery behaviours to shed greater light on how they develop, and how to foster them.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The research reported here was supported by R305A150142 from the Institute for Education Sciences, and by Character Lab.

References

Abrahams, L., Pancorbo, G., Primi, R., Santos, D., Kyllonen, P., John, O. P., & De Fruyt, F. (2019). Social-emotional skill assessment in children and adolescents: Advances and challenges in personality, clinical, and educational contexts. *Psychological Assessment, 31*(4), 460–473. https://dx.doi.org/10.1037/pas0000591

Arendasy, M., Sommer, M., Herle, M., Schützhofer, B., & Inwanschitz, D. (2011). Modeling effects of faking on an objective personality test. *Journal of Individual Differences, 32*, 210–218. https://dx.doi.org/10.1027/1614-0001/a000053

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*, 596–612. https://dx.doi.org/10.1037/0022-3514.63.4.596

Bettinger, E., Ludvigsen, S., Rege, M., Solli, I. F., & Yeager, D. (2018). Increasing perseverance in math: Evidence from a field experiment in Norway. *Journal of Economic Behavior & Organization, 146*, 1–15. https://dx.doi.org/10.1016/j.jebo.2017.11.032

Blackwell, L., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*(1), 246–263. https://dx.doi.org/10.1111/j.1467-8624.2007.00995.x

Brown, E. D. (2009). Persistence in the face of academic challenge for economically disadvantaged preschool children. *Journal of Early Childhood Research, 7*(2), 175–186. https://dx.doi.org/10.1177/1476718X09102650

Burnette, J. L., O'Boyle, E. H., VanEpps, E. M., Pollack, J. M., & Finkel, E. J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin, 139*(3), 655–701. https://dx.doi.org/10.1037/a0029531

Caspi, A., Elder, G. H., & Bem, D. J. (1988). Moving away from the world: Life-course patterns of shy children. *Developmental Psychology*, *24*(6), 824–831. https://dx.doi.org/10.1037/0012-1649.24.6.824

Choudhury, N., & Gorman, K. S. (1999). The relation between reaction time and psychometric intelligence in a rural Guatemalan adolescent population. *International Journal of Psychology*, *34*(4), 209–217. https://dx.doi.org/10.1080/002075999399855

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment, 31*(12), 1412–1427. https://dx.doi.org/10.1037/pas0000626

Cohen, P., Cohen, J., Aiken, L. S., & West, S. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research. 34*(3), 315–346. https://dx.doi.org/10.1207/S15327906MBR3403_2

Costenbader, V., & Ngari, S. M. (2001). A Kenyan standardization of the Raven's coloured progressive matrices. *School Psychology International, 22*(3), 258–268. https://dx.doi.org/10.1177/0143034301223003

Cutumisu, M., Blair, K. P., Chin, B., & Schwartz, D. L. (2015). Posterlet: A game-based assessment of children's choices to seek feedback and to revise. *Journal of Learning Analytics, 2*(1), 49–71. https://dx.doi.org/10.18608/jla.2015.21.4

Diener, C. I., & Dweck, C. S. (1978). An analysis of learned helplessness: Continuous changes in performance, strategy, and achievement cognitions following failure. *Journal of Personality and Social Psychology, 36*(5), 451–462. https://dx.doi.org/10.1037/0022-3514.36.5.451

Diener, C. I., & Dweck, C. S. (1980). An analysis of learned helplessness: II. The processing of success. *Journal of Personality and Social Psychology, 39*(5), 940–952. https://dx.doi.org/10.1037/0022-3514.39.5.940

Diener, E., Ingelhart, R. F., & Tay, L. (2013). Theory and validity of life satisfaction scales. *Social Indicators Research, 112*(5), 497–527. https://dx.doi.org/10.1007/s11205-012-0076-y

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality, 45*(3), 259–268. https://dx.doi.org/10.1016/j.jrp.2011.02.004

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*(4), 237–251. https://doi.org/10.3102/0013189X15584327

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development.* New York: Psychology Press, Taylor & Francis Group. https://doi.org/10.4324/9781315783048

Dweck, C. S., Chiu, C., & Hong, Y. (1995). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry*, *6*(4), 267–285. https://dx.doi.org/10.1207/s15327965pli0604_1

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256–273. https://dx.doi.org/10.1037/0033-295X.95.2.256

Eisinga, R., te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Person, Cronbach, or Spearman-Brown? *International Journal of Public Health, 58*(4), 637–642. https://dx.doi.org/10.1007/s00038-012-0416-3

Elliot, A., McGregor, H., & Gable, S. (1999). Achievement goals, study strategies, and exam performance: A mediational analysis. *Journal of Educational Psychology, 91,* 549–563. https://dx.doi.org/10.1037/0022-0663.91.3.549

Enkavi, Z. A, Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences of the United States of America, 116*(12), 5472–5477. https://dx.doi.org/10.1073/pnas.1818430116

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect sizes in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156–168. https://dx.doi.org/10.1177/2515245919847202

Galla, B. M., Plummer, B. D., White, R. E., Meketon, D., D'Mello, S. K., & Duckworth, A. L. (2014). The academic diligence task (ADT): Assessing individual differences in effort on tedious but important schoolwork. *Contemporary Education Psychology, 39*(4), 314–325. https://dx.doi.org/10.1016/j.cedpsych.2014.08.001

Gibson, A., Kitto, K., & Bruza, P. (2016). Towards the discovery of learner metacognition from reflective writing. *Journal of Learning Analytics*, *3*(2), 22–36. https://dx.doi.org/10.18608/jla.2016.32.3

Gottfried, A. E. (1985). Academic intrinsic motivation in elementary and junior high school students. *Journal of Educational Psychology, 77*(6), 631–645. https://dx.doi.org/10.1037/0022-0663.77.6.631

Haimovitz, K., & Dweck, C. S. (2016). Parents' views of failure predict children's fixed and growth intelligence mind-sets. *Psychological Science, 27*(6), 859–869. https://dx.doi.org/10.1177/0956797616639727

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavioral Research Methods, 50*(3), 1166–1186. https://dx.doi.org/10.3758/s13428-017-0935-1

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2-3), 61–83. https://dx.doi.org/10.1017/S0140525X0999152X

Joksimović, S., Gašević, D., Loughin, T. M., Kovanović, V., & Hatala, M. (2015). Learning at distance: Effects of interaction traces on academic achievement. *Computers & Education, 87*, 204–217. https://dx.doi.org/10.1016/j.compedu.2015.07.002

Knafo, A., Israel, S., Darvasi, A., Bachner-Melman, R., Uzefovsky, F., Cohen, L., …, Ebstein, R. P. (2007). Individual differences in allocation of funds in the dictator game associated with length of the arginine vasopressin 1a receptor RS3 promoter region and correlation between RS3 length and hippocampal mRNA. *Genes Brain Behavior, 7*(3), 266–275. https://dx.doi.org/10.1111/j.1601-183X.2007.00341.x

Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., & Hatala, M. (2015). Penetrating the black box of time-on-task estimation, *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (LAK '15)*, 16–20 March 2015, Poughkeepsie, NY (pp. 184–193). New York: ACM. https://dx.doi.org/10.1145/2723576.2723623

Kraft, M. A. (2019). Interpreting effect sizes of education interventions. *EdWorkingPaper No. 19-10*. https://dx.doi.org/10.26300/8pjp-2z74

Krumm, A. E., Beattie, R., Takahashi, S., D'Angelo, C., Feng, M., & Cheng, B. (2016). Practical measurement and productive persistence: Strategies for using digital learning system data to drive improvement. *Journal of Learning Analytics, 3*(2), 116–138. https://dx.doi.org/10.18608/jla.2016.32.6

Licht, B. G., & Dweck, C. S. (1984). Determinants of academic achievement: The interaction of children's achievement orientations with skill area. *Developmental Psychology, 20*(4), 628–636. https://dx.doi.org/10.1037/0012-1649.20.4.628

Lucas, R. E., & Donnellan, B. M. (2012). Estimating the reliability of single-item life satisfaction measures: Results from four national panel studies. *Social Indicators Research, 105*(3), 323–331. https://dx.doi.org/10.1007/s11205-011-9783-z

Mangels, J. A., Butterfield, B., Lamb, J., Good, C., & Dweck, C. (2006). Why do beliefs about intelligence influence learning success? A social cognitive neuroscience model. *Social Cognitive and Affective Neuroscience, 1*(2), 75–86. https://dx.doi.org/10.1093/scan/nsl013

Marcus, B. (2006). Relationships between faking, validity, and decision criteria in personnel selection. *Psychology Science, 48*(3), 226–246. https://psycnet.apa.org/record/2006-21801-003

Martin, C. K., Nacu, D., & Pinkard, N. (2016). Revealing opportunities for 21st century learning: An approach to interpreting user trace log data. *Journal of Learning Analytics*, *3*(2), 37–87. https://dx.doi.org/10.18608/jla.2016.32.4

Meindl, P., Yu, A., Galla, B., Quirk, A., Haeck, C., Parker, G. J., …, Duckworth, A. L. (2019). A brief behavioral measure of frustration tolerance predicts academic achievement immediately and two years later. *Emotion, 19*(6), 1081–1092. https://dx.doi.org/10.1037/emo0000492

Miller, R., Greene, B., Montalvo, G., Ravindran, B., & Nichols, J. (1996). Engagement in academic work: The role of learning goals, future consequences, pleasing others, and perceived ability. *Contemporary Educational Psychology, 21*, 388–422. https://dx.doi.org/10.1006/ceps.1996.0028

Mischel, W., Shoda, Y., & Peake, P. K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology, 54*, 687–696. https://dx.doi.org/10.1037/0022-3514.54.4.687

Moser, J. S., Schroder, H. S., Heeter, C., Moran, T. P., & Lee, Y. (2011). Mind your errors: Evidence for a neural mechanism linking growth mind-set to adaptive posterror adjustments. *Psychological Science, 22*(12), 1484–1489. https://dx.doi.org/10.1177/0956797611419520

Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology, 75*(1), 33–52. https://dx.doi.org/10.1037/0022-3514.75.1.33

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales.* Oxford, U.K.: Oxford Psychologists Press.

Robins, R. W., & Pals, J. L. (2002). Implicit self-theories in the academic domain: Implications for goal orientation, attributions, affect, and self-esteem change. *Self and Identity, 1*(4), 313–336. https://dx.doi.org/10.1080/15298860290106805

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380–1400. https://dx.doi.org/10.1177/0002764213498851

Soland, J., Zamarro, G., Cheng, A., & Hitt, C. (2019). Identifying naturally occurring direct assessments of social-emotional competencies: The promise and limitations of survey and assessment disengagement metadata. *Educational Researcher, 48*(7), 466–478. https://dx.doi.org/10.3102/0013189X19861356

Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology, 95*(5), 1202–1216. https://dx.doi.org/10.1037/a0013314

Ventura, M., & Shute, V. (2013). The validity of a game-based assessment of persistence. *Computers in Human Behavior, 29*, 2568–2572. https://dx.doi.org/10.1016/j.chb.2013.06.033

Ventura, M., Shute, V., & Zhao, W. (2013). The relationship between video game use and a performance-based measure of persistence. *Computers & Education, 60*(1), 52–58. https://dx.doi.org/10.1016/j.compedu.2012.07.003

Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology, 92*(1), 82–96. https://dx.doi.org/10.1037/0022-3514.92.1.82

Willerman, L., Turner, R. G., & Peterson, M. (1976). A comparison of the predictive validity of typical and maximal personality measures. *Journal of Research in Personality, 10*(4), 482–492. https://dx.doi.org/10.1016/0092-6566(76)90063-5

Yeager, D. S., & Bryk, A. (2015). *Practical measurement.* Unpublished manuscript, University of Texas at Austin.

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., …, Dweck, C. S. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology, 108*(3), 374–391. https://dx.doi.org/10.1037/edu0000098

Ziegler, M., Schmidt-Atzert, L., Bühner, M., & Krumm, S. (2007). Fakability of different measurement methods for achievement motivation: Questionnaire, semi-projective, and objective. *Psychology Science, 49*, 291–307. https://psycnet.apa.org/record/2008-02884-001

