

RESEARCH

Open Access



# Quantifying fear of failure in STEM: modifying and evaluating the Performance Failure Appraisal Inventory (PFAI) for use with STEM undergraduates

Meredith A. Henry<sup>1</sup>, Shayla Shorter<sup>1</sup>, Louise K. Charkoudian<sup>2</sup>, Jennifer M. Heemstra<sup>1</sup>, Benjamin Le<sup>3</sup> and Lisa A. Corwin<sup>4\*</sup> 

## Abstract

**Background:** The ability to navigate obstacles and embrace iteration following failure is a hallmark of a scientific disposition and is hypothesized to increase students' persistence in science, technology, engineering, and mathematics (STEM). However, this ability is often not explicitly explored or addressed by STEM instructors. Recent collective interest brought together STEM instructors, psychologists, and education researchers through the National Science Foundation (NSF) research collaborative *Factors affecting Learning, Attitudes, and Mindsets in Education network* (FLAMEnet) to investigate intrapersonal elements (e.g., individual differences, affect, motivation) that may influence students' STEM persistence. One such element is *fear of failure* (FF), a complex interplay of emotion and cognition occurring when a student believes they may not be able to meet the needs of an achievement context. A validated measure for assessing FF, the Performance Failure Appraisal Inventory (PFAI) exists in the psychological literature. However, this measure was validated in community, athletic, and general undergraduate samples, which may not accurately reflect the motivations, experiences, and diversity of undergraduate STEM students. Given the potential role of FF in STEM student persistence and motivation, we felt it important to determine if this measure accurately assessed FF for STEM undergraduates, and if not, how we could improve upon or adapt it for this purpose.

**Results:** Using exploratory and confirmatory factor analysis and cognitive interviews, we re-validated the PFAI with a sample of undergraduates enrolled in STEM courses, primarily introductory biology and chemistry. Results indicate that a modified 15-item four-factor structure is more appropriate for assessing levels of FF in STEM students, particularly among those from groups underrepresented in STEM.

**Conclusions:** In addition to presenting an alternate factor structure, our data suggest that using the original form of the PFAI measure may significantly misrepresent levels of FF in the STEM context. This paper details our collaborative validation process and discusses implications of the results for choosing, using, and interpreting psychological assessment tools within STEM undergraduate populations.

**Keywords:** Fear of failure, Assessment, Measure validation, Undergraduate STEM education, Academic failure, Academic challenge

\* Correspondence: [lisa.corwin@colorado.edu](mailto:lisa.corwin@colorado.edu)

<sup>4</sup>Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, 1900 Pleasant St., 334 UCB, Boulder, CO 80309, USA  
Full list of author information is available at the end of the article

## Introduction

The ability to navigate obstacles and embrace an iterative process in response to failure is considered a hallmark of the scientific disposition and has been hypothesized to increase students' persistence in STEM (Harsh et al. 2011; Laursen et al. 2010; Lopatto et al. 2008; Simpson and Maltese 2017; Thiry et al. 2012). The ways in which this ability can be fostered through undergraduate science, technology, engineering, or mathematics (STEM) education is a topic both historically underexplored by researchers and under-addressed by explicit instructor-driven curricula (Simpson and Maltese 2017; Traphagen 2015). However, recent increased interest in investigating the effects of various intrapersonal attributes on STEM students' ability to navigate scientific obstacles has set the stage for promising educational research in this arena. *Intrapersonal elements*, previously referred to more broadly as "noncognitive factors" (Henry et al. 2019), include a subset of one's individual competencies not related to one's intelligence or knowledge. For example, intrapersonal elements include mindsets, attitudes, and beliefs, among other elements related to one's understanding of their own experience (this is contrasted with *interpersonal elements*, such as empathy and other social skills, which involve recognizing *others'* perspectives and experiences; Farrington 2019; National Research Council 2012). Many of these intrapersonal elements—such as fear of failure (FF), the topic of this work—are predicted to influence students' engagement with challenges, responses to failure, and subsequent academic success (Henry et al. 2019). Yet, unlike some predictors of success that can be measured directly (e.g., prior achievement), most intrapersonal elements consist of latent variables, which cannot be directly observed or measured. Such variables must be assessed using multiple metrics (often multiple questions on a survey) that together allow us to estimate levels of the underlying construct (Knekta et al. 2019). Unfortunately, measures for these elements often either do not exist or may not be valid for our population of interest, STEM undergraduates. This is the case for FF. In this study, we build on prior work that describes the validity of a measure of FF, the Performance Failure Appraisal Inventory (PFAI). We investigate the validity of this instrument and work to improve and modify it for use with STEM undergraduates. Our aim was to provide a suitable revision of the PFAI and make it available to instructors and education researchers to measure FF in undergraduate STEM contexts.

### Fear of failure

FF involves a complex interplay of emotion (Martin and Marsh 2003), personality (Noguera et al. 2013), and cognition (Conroy 2001). Historically, research has focused

separately on these three aspects. Past researchers have described FF as either (a) purely affective, consisting of feelings of anxiety, nervousness, or worry when considering future failures; (b) an aspect of personality, for example having a high degree of neuroticism that consistently contributes to FF across all contexts; or (c) a context-specific cognitive assessment that evaluates a given situation as a threat to success (e.g., evaluation of failing a class as being a determinant of admission to medical school, and given this, fear of failing). However, more recent work recognizes that all of these domains are interrelated and contribute to the most comprehensive explanation of FF (Henry et al. 2019). Specifically, Cacciotti (2015) defined FF as a "temporary cognitive and emotional reaction towards environmental stimuli that are apprehended as threats in achievement contexts" (p. 59). An achievement context includes any situation in which (1) some task must be performed, (2) the task will be evaluated against standards or expectations, and (3) one must have certain competencies in order to carry out the task to those standards (Cacciotti 2015). In other words, FF is manifested in anxiety-based thoughts and emotions when one believes they may be unable to meet the demands of an achievement context. It is important to distinguish this multidimensional view of FF from constructs that solely describe emotion, such as anxiety. While these have been used as analogs for FF in the past, our modern understanding of FF recognizes that focusing only on the emotional aspects of the experience provides an incomplete understanding. For example, focusing only on emotion fails to recognize the cognitive appraisals of an achievement context that are often the root cause of affective states and may constitute specific targets for interventions seeking to alleviate FF. In other words, exploring the emotions related to FF (like anxiety) is necessary, but not sufficient, for a complete understanding of FF (Henry et al. 2019).

A key component of this definition is that one's level of FF may change based upon the specific details of the achievement context or other contextual factors (Conroy 2001). For example, if two students enrolled in introductory biology have different goals, they will create different achievement contexts based on those expectations. If Student A is only enrolled because the course fulfills a general education requirement, the demands to satisfy the achievement context may be relatively low (e.g., just pass the class). The student is therefore less likely to experience FF. However, Student B, who is pursuing graduate study or a health career, is likely to judge the achievement context to be much more demanding (e.g., anything less than an A is unacceptable). While these students may generally differ to some extent in their baseline levels of FF, Student B will also likely experience greater FF, in part because they perceive the stakes of failure to be higher.

FF is broadly recognized as an element that can lead to avoidance of challenge, lower motivation, and self-impeding behaviors (e.g., making excuses, reduced effort, etc.; Chen et al. 2009). It has been studied extensively in K-12 contexts (e.g., Caraway et al. 2003; De Castella et al. 2013; Pelin and Subasi 2020) and in certain nonacademic contexts, such as entrepreneurship (e.g., Cacciotti et al. 2016) and sports (e.g., Conroy et al., 2001; Sagar et al., 2010). It has also been studied extensively in undergraduate students broadly (Bartels and Herman 2011; Bledsoe and Baskin 2014; Chen et al. 2009; Elliot and Church 1997; Elliot and Church 2003; Elliot and McGregor 2001; Elliot and Thrash 2004). Yet, despite its potential to impact achievement, FF has not been studied extensively in STEM undergraduate contexts. This is surprising given that “embracing failure” is seen as a necessary skill for professional scientists and that STEM individuals are known to view failures in ways distinct from those in other fields (Simpson and Maltese 2017). Although studies are few in number, several lines of evidence suggest that STEM students experience FF and that this may either limit engagement in STEM learning or, in some cases, prevent engagement altogether. Researchers have found that FF positively predicts procrastination behaviors for pre-health undergraduates (Zhang et al., 2018) and that this relationship extends to STEM graduate students in statistics courses (Onwuegbuzie, 2004). Similarly, work on understanding the causes of student anxiety during active learning in STEM classrooms cites a closely related construct, fear of negative evaluation by others, as an important cause of anxiety that can hamper students’ motivation to participate in class (Cooper et al. 2018; Downing et al. 2020). Ceyhan and Tillotson (2020) found that undergraduate STEM majors weighed their FF as an element influencing their motivation to engage in undergraduate research, citing the emotional cost of engagement. This is especially notable considering the enthusiasm for and movement towards research-based courses that are more likely to expose students to scientific failures (Auchincloss et al. 2014; Corwin et al. 2015; Gin et al. 2018). Indeed, FF may become more salient to students engaged with these new pedagogies. Prior work also suggests that FF differs among male- and female-identified STEM undergraduate students (Nelson et al., 2021), suggesting that we need to consider differential effects of FF across identities in STEM. Finally, and most importantly, FF may predict whether or not students ultimately choose a STEM major or choose to remain a STEM major after their first semester at college (Nelson et al. 2019). FF may contribute to the extensive drop in STEM majors typically seen after the first year in STEM fields (Nelson et al. 2019; Seymour and Hunter 2019). STEM instructors recognize that assisting students in coping with

failure and alleviating FF are important priorities when training future scientists (Gin et al. 2018; Henry et al. 2019; Simpson and Maltese 2017). However, we must be able to accurately measure the effects of these efforts if we are to understand how teaching practices can serve to alleviate FF.

### Measuring fear of failure

Given the complex nature of FF, it can be difficult to conceptualize a valid measure which fully captures all of its properties. For example, risk aversion was historically used as a proxy for FF, but contemporary researchers now acknowledge that, while risk aversion may tap into some of the emotional aspects of FF, it likely does not fully represent the personality or cognitive aspects (Noguera et al. 2013), nor is it responsive to changes in context (Conroy 2001). Beginning in 2001, Conroy and colleagues addressed these concerns by attempting to understand the causes of FF at the individual level, resulting in the creation and refinement of a multidimensional assessment measure: the PFAI.

To characterize FF, Conroy et al. (2001) conducted interviews with eight adult elite athletes and eight adult elite performing artists (50% female). In these interviews, subjects provided insight into their definitions of failure, what situations or contexts they considered “failures” in the past, and how they reacted to experiencing those situations (Conroy et al. 2001). Based on a content analysis of these interviews, Conroy et al. (2001) created 89 items that could be classified under ten broad sources of FF (e.g., fear of an uncertain future). They then asked 396 high school and college-aged student-athletes (mean age = 19.3 years, SD = 4.3) to respond to these 89 items. Each statement evoked a situation in which the student was “failing” or “not succeeding” and students rated each item on a scale of “Do not believe [this to be true] at all (– 2)” to “Believe [this to be true] 100% of the time (+ 2)”. For example, students read the statement “When I am failing, my future seems uncertain” and then selected whether they believed this to be true and to what degree. Conroy and colleagues then used factor analysis, which groups items according to statistical relationships that correspond to psychological constructs—in this case, types of FF. This work narrowed the number of PFAI items down to 41, loading strongly onto five factors (meaning that the items cluster broadly into five meaningful categories or *dimensions*, rather than the originally proposed 10, see Knekta et al. (2019) for an explanation of how factors are formed). Subsequent factor analyses with samples of only college student participants further reduced the number of items on the PFAI from 41 to 25, with five items measuring each dimension, or reason for demonstrating FF (Table 1; Conroy et al. 2002).

**Table 1** Sample items representing the five dimensions of the original PFAI (Conroy et al. 2002)

Dimension/factor name	Abbreviation	Sample item
Fear of shame or embarrassment	FSE	"When I am not succeeding, I worry about what others think of me."
Fear of devaluing one's self-estimate	FDSE	"When I am failing, I blame my lack of talent."
Fear of having an uncertain future	FUF	"When I am failing, it upsets my 'plan' for the future."
Fear of important others losing interest	FIOLI	"When I am not succeeding, some people are not interested in me anymore."
Fear of upsetting important others	FUIO	"When I am failing, I lose the trust of people who are important to me."

### The current work

Addressing FF is especially important when considering the broader challenge of finding and validating measures of intrapersonal elements for STEM undergraduates (Knekta et al. 2019; Rowland et al. 2019) and the more specific challenge of assessing the influence of intrapersonal elements on students' resilience, motivation to engage in challenging tasks, and ability to navigate obstacles when they arise (Henry et al. 2019). Previous research utilizing the PFAI as an assessment measure has found that increased FF is related to reduced challenge engagement (Bledsoe and Baskin 2014). This can most clearly be seen in high-FF students who demonstrate self-impeding behaviors by reducing effort or making excuses before failure occurs, thereby protecting their self-worth in the short term at the risk of long-term success (Berglas and Jones 1978; Chen et al. 2009; Cox 2009; Zuckerman and Tsai 2005). While these results were found in academic, community, and broad college contexts, they have not yet been investigated specifically in undergraduate STEM education, a context in which challenge engagement is likely critical for progress (Henry et al. 2019) but where failure is also a commonly accepted part of the process (Simpson and Maltese 2017). The contextual nature of both FF (Cacciotti 2015) and human cognitive appraisals (Schwarz 1999) suggests that FF is likely to manifest in significantly different ways depending upon the achievement context(s) one is assessing. Therefore, college students in a STEM context are likely to experience FF differently than students in high school or non-STEM courses. Many students enrolled in STEM courses enter with intentions of pursuing graduate study or health careers (e.g., Gasiewski et al. 2012), which can make achievement contexts more salient. In addition, the active research that students may engage in during the course of STEM education often includes tasks that have a higher likelihood of failure or not achieving a stated research goal (Auchincloss et al. 2014), and this may represent one of the first times students encounter an academic situation in which judging success against the achievement context is difficult or unclear. All of these contextual factors influence how STEM students experience FF and also how they will respond to assessments of FF. To investigate and

understand FF in STEM contexts, we need to ensure that we can accurately measure FF for STEM students.

Any education research conducted on a topic will only be as strong as the assessment tools used for the constructs of interest (Cronbach and Meehl 1955). While the PFAI was originally constructed using a sample that included some college students, these students were only used for factor analyses to reduce the number of items and refine the measure for certain types of FF. They were not interviewed as part of the initial creation of the items or to ascertain how they interpreted the items and if this interpretation matched the assumptions and definitions of FF researchers. This reveals a critical unmet need, because there is much about college contexts—and undergraduate STEM achievement contexts more specifically—which could affect students' responses and/or alter response patterns to the PFAI. Such contextual factors could make Conroy's proposed factor structure inappropriate and invalid for these specialized populations of interest. This is important, because if the PFAI is not valid for undergraduate STEM samples, using it for education research could lead to misrepresentation of FF levels and faulty conclusions about levels of FF present in STEM classrooms or the efficacy of interventions that aim to reduce FF within that context. As such, we set out to build upon the work of Conroy et al. (2002) to (a) ascertain specifically how STEM undergraduates interpret PFAI questions related to Conroy et al.'s (2002) proposed dimensions of FF and (b) refine the PFAI for assessing FF in undergraduate STEM contexts. First, we used confirmatory factor analysis (CFA) to test whether Conroy's proposed five-factor model is appropriate for measuring STEM undergraduates' FF. Then, we used exploratory factor analysis (EFA) to consider alternative factor structures and evaluate whether any offer a better fit for data from an undergraduate STEM sample. Once we identified the "strongest" factor structure for the initial sample of STEM students, we performed additional CFA analyses in several other samples to confirm this structure. We then asked whether that factor structure remains a good fit for science PEERs—persons excluded because of their ethnicity or race (Asai 2020) using an additional CFA. Finally, we conducted a series of cognitive interviews among

students to assess the face and content validity of our final revised measure to (a) ensure that students find all items clear and easy to understand and (b) better characterize the nuance with which STEM undergraduate students interpret the PFAI items. The progression of our analyses and their basic goals is outlined in Fig. 1, with more details about each analysis' sample and the location of specific results detailed in Table 2. Below, we describe each of these steps, their methods, results, and a brief discussion of their findings before concluding with a broad discussion of our modified measure and its suitability for assessing FF in undergraduate STEM student populations.

### Step 1: Confirmatory factor analyses (CFAs) of existing models

CFAs serve to test whether a measure of a construct—in this case the PFAI—is consistent with the proposed understanding of that construct (i.e., FF) and its components. A good “fit” of the data collected with a particular measure to the proposed conceptual model indicates that the measure is accurate with regard to the researchers' understanding of the construct and its components. When there is strong a priori and/or empirical evidence supporting a conceptual model—as in the case of the PFAI—it is best to start with CFAs when exploring measure utility for new populations (Knetaka et al., 2019). Thus, the purpose of our initial CFA analyses was twofold: (a) to investigate whether data collected from a sample of undergraduate STEM students fit to the current factor structure (twenty-five items on five factors) proposed by Conroy et al. (2002) and (b) to assess whether or not a change in item wording to prompt students to consider struggles and failures *specifically in STEM contexts* improves the model fit. We reasoned that the alternate wording would result in an improved model fit, as it addresses the specific, unique (as discussed above) STEM education context in which students find themselves frequently confronting academic challenges and failures. We explored this possibility with the aim of creating a version of the PFAI best suited to assess STEM-related FF in undergraduate students.

### Methods

#### Participants

Four hundred and twenty-three undergraduate students were recruited for this study during Spring/Summer 2018. Students were recruited with the aid of STEM instructors at a diverse group of institutions—public and private; rural and urban; liberal arts and research-intensive—across multiple regions of the USA. Recruiting instructors were members of FLAMENet, an NSF-funded research collaborative which brings together diverse STEM instructors, education researchers, and

social scientists to conduct research and create resources aimed at fostering the next generation of resilient and innovative scientists (<https://qubeshub.org/community/groups/flamenet/>). Instructors announced the research opportunity to students either during class, via the course learning management system, via email, or on social media. All recruited students were enrolled in a STEM course at the time of the study. Two hundred and thirty-five students in this volunteer sample provided complete surveys and were included in the final data set. Students in this sample predominantly identified as female (68.1%) and Caucasian (81.7%), with a majority describing themselves as STEM majors (90%). A full breakdown of participant characteristics can be found in Table 3, column 3.

#### Instruments

**Fear of failure** The 25-item version of the PFAI was employed (Conroy et al. 2002; the original measure can be viewed in Table S1). Using this measure, we asked participants to endorse certain beliefs regarding the likely consequence(s) of failure on a scale of 1 (“I believe this is never true of me”) to 5 (“I believe this is true of me all of the time”). This scale was modified from the original 5-point scale of – 2 “Do not believe at all” to + 2 “Believe 100% of the time” based on anecdotal preliminary feedback from undergraduate research assistants that the modified response scale is easier to understand. Evidence of the validity and reliability of this 25-item version of the PFAI (Conroy 2001) has previously been gathered in a general college sample. In that study, Conroy et al. (2002) produced a final well-fitting model that accounted for the five proposed dimensions—fear of shame or embarrassment (FSE), fear of devaluing one's self-estimate (FDSE), fear of having an uncertain future (FUF), fear of important others losing interest (FIOLI), and fear of upsetting important others (FUIO)—each of which consists of a group of five questions that corresponds to the specified dimension, along with the higher-order overall FF that can be derived from averaging these five subscales. (From here forward, we refer to these groups of questions as “dimensions” since they are inseparable from the FF dimensions. When discussing the mathematical fit of items to these dimensions resulting from factor analysis, we use the term “factors”.) Fit statistics for this original model are provided at the top of Table 4. Internal reliability for this form of the PFAI has previously been demonstrated to be high, with Cronbach's alpha of all twenty-five items at 0.91 and for the higher-order FF dimension derived by the mean of the five factors at 0.82 (Conroy et al. 2002; Cronbach's alpha is a statistical measure which assesses the degree to which items on a scale are correlated with each other,

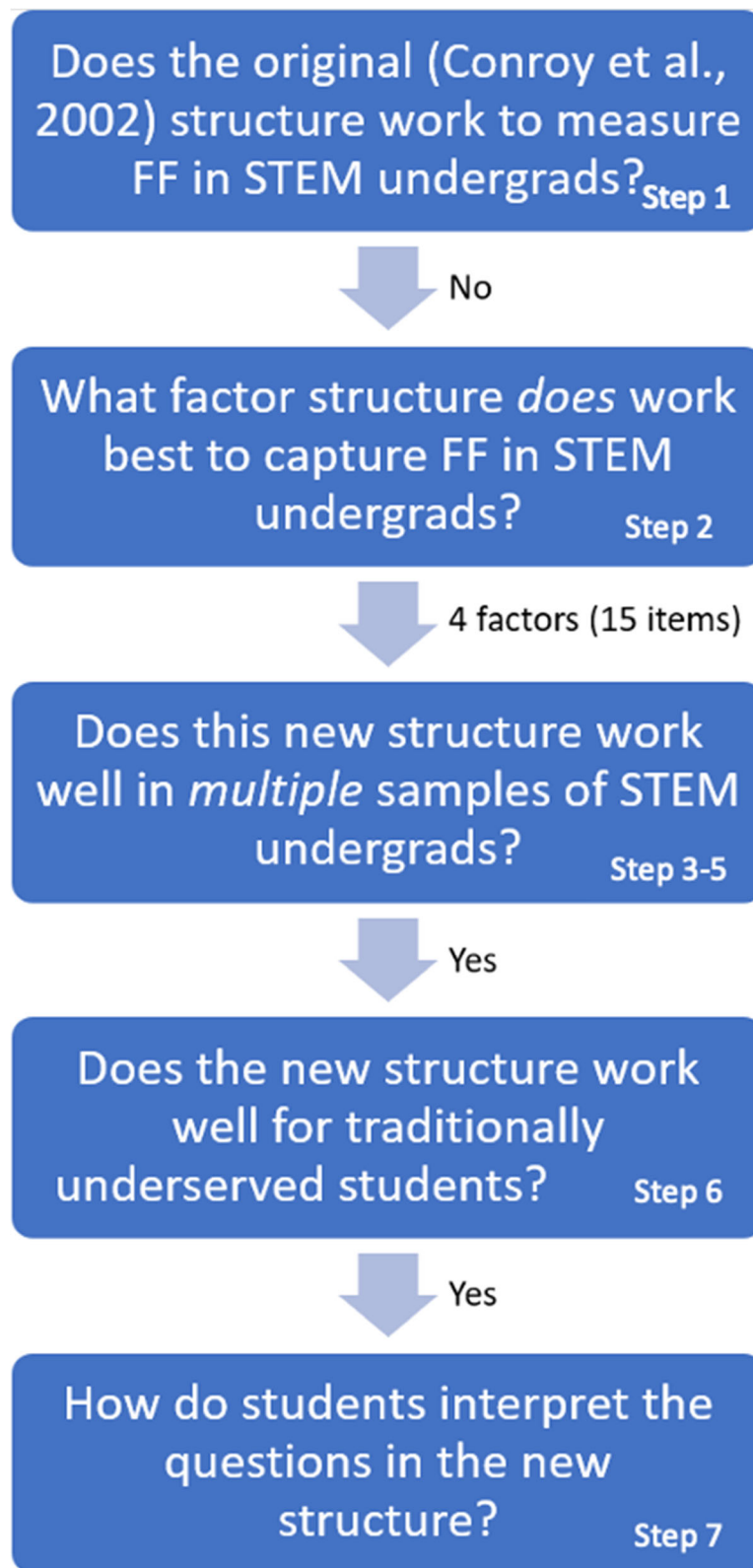


Fig. 1 Progression of research steps

**Table 2** Order of tests

	<i>Sample</i>			
<i>Analyses</i>	Recruited for validation study, Summer 2018 ( <i>n</i> = 235)	Pulled from existing intervention study, Fall 2018 ( <i>n</i> = 1309)	Pulled from existing intervention study, Fall 2019 ( <i>n</i> = 433)	PEER students across timepoints, Summer 2018, Fall 2018, Fall 2019 ( <i>n</i> = 280)
Step 1	CFA to test existing PFAI structure (Conroy et al. 2002)			
	<ul style="list-style-type: none"> <li>• Table 3, Column 3 (Participants)</li> <li>• Table 4 (Results)</li> </ul>			
Step 2	EFA to define new factor structure			
	<ul style="list-style-type: none"> <li>• Table 3, Column 4 (Participants)</li> <li>• Table 5, Rows 3-5 (Results)</li> </ul>			
Step 3	CFA to confirm revised structure			
	<ul style="list-style-type: none"> <li>• Table 3, Column 3 (Participants)</li> <li>• Table 5, Row 7 (Results)</li> </ul>			
Step 4	CFA to confirm revised structure in new dataset			
	<ul style="list-style-type: none"> <li>• Table 3, Column 5 (Participants)</li> <li>• Table 5, Row 8 (Results)</li> </ul>			
Step 5	CFA to confirm revised structure in original data set			
	<ul style="list-style-type: none"> <li>• Table 3, Column 3 (Participants)</li> <li>• Table 5, Row 9 (Results)</li> </ul>			
Step 6	CFA to test fit of revised structure among PEER undergraduates			
	<ul style="list-style-type: none"> <li>• Table 3, Column 6 (Participants)</li> <li>• Table 5, Row 10 (Results)</li> </ul>			
Step 7	Cognitive interviews to assess face validity			
	<ul style="list-style-type: none"> <li>• Table 7 (Participants)</li> </ul>			

with values closer to one indicating stronger relationships. If items on a scale do, in fact, all measure the same construct, one would expect to see high consistency as measured by Cronbach's alpha; Cronbach (1951). In this study, PFAI items were framed in two ways. In the "general" condition, the PFAI items were introduced by asking students to "consider the way you feel and act when you face failures and challenges." The PFAI items themselves were introduced with the following: "For the following questions, please consider challenges and failures that you face in general. For each question, indicate how often you believe each statement is true of you." By contrast, our "STEM" condition introduced this section of the survey by asking respondents to "consider the way you feel and act when you face failures and challenges in your STEM courses". And the questions themselves were introduced by reminding students to "please consider challenges and failures that you face specifically in your STEM course(s). For each question, indicate how often you believe each

statement is true of you." Students also provided qualitative responses to a set of questions that asked them to describe a recent time when they experienced a failure or challenge (again, either in general or in a STEM context specifically) and how upsetting they found this event on a scale from 0 to 10, with 10 being the most upsetting. These qualitative items are discussed more in *Step 7* below.

**STEM anxiety** To quantify the overall level of anxiety surrounding the academic context of STEM courses, we asked students the following: "On a scale from 0 to 10, how anxious are you about your performance in your STEM classes?" Here, 0 indicated "not at all anxious" while 10 indicated "extremely anxious." Students utilized the entire range of possible responses for this question, with a mean response of 6.37 (SE 0.175; variance 6.670). Cognitive interviews (see *Step 7*) were used to assess the face validity of this question, with students indicating that they found it simple and straightforward.

**Table 3** Demographic characteristics of participants

Variable	Value	Steps 1 and 5: Initial CFA sample (Summer 2018; N = 235)	Steps 2 and 3 : EFA sample (Fall 2018; N = 1309)	Step 4: Second CFA sample (Fall 2019; N = 433)	Step 6: PEER CFA sample (Summer 2018, Fall 2018, Fall 2019; N = 280)
		Frequency (%)	Frequency (%)	Frequency (%)	Frequency (%)
Gender	Female	160 (68.1)	888 (68.4)	272 (62.8)	195 (69.6)
	Male	52 (22.1)	384 (29.6)	150 (34.6)	81 (28.9)
	Non-binary	3 (1.3)	4 (0.3)	4 (0.9)	3 (1.1)
	Prefer not to answer	2 (0.9)	2 (0.2)	6 (1.4)	1 (0.4)
Major <sup>1</sup>	Biology	103 (43.8)	456 (36.05)	-	40 (22.1)
	Chemistry	39 (16.6)	100 (7.91)	-	15 (22.1)
	Engineering	15 (6)	8 (0.63)	-	2 (1.1)
	Env. Science	10 (4)	4 (0.32)	-	2 (1.1)
	Pre-health	12 (5)	101 (7.98)	-	40 (22.1)
	Kinesiology	11 (4.7)	20 (1.58)	-	-
	Mathematics	3 (1.3)	54 (4.27)	-	1 (0.05)
	Neuroscience	15 (6)	89 (7.04)	-	20 (11.05)
	Physics	7 (3)	135 (10.67)	-	2 (1.1)
	Psychology	14 (6)	26 (2.06)	-	10 (5.52)
	Other	5 (2)	64 (5.06)	-	25 (13.81)
	Undeclared	1 (0.4)	208 (16.4)	-	24 (13.26)
Age	18–20	191 (81.3)	1184 (90.4)	320 (73.9)	227 (81.1)
	21–23	37 (15.7)	52 (3.9)	58 (13.4)	26 (9.3)
	24 and older	7 (3.0)	64 (4.9)	55 (12.7)	24 (8.6)
Class standing	First year	68 (28.9)	306 (23.6)	184 (42.5)	112 (40.0)
	Second year	63 (26.8)	302 (23.3)	135 (31.2)	112 (40.0)
	Third year	41 (17.4)	60 (4.6)	50 (11.5)	31 (11.1)
	Fourth year	41 (17.4)	36 (2.8)	31 (7.2)	15 (5.4)
	Other	4 (1.7)	6 (0.5)	32 (7.4)	10 (3.6)
Race	American Indian	1 (0.4)	6 (0.5)	3 (0.7)	62 (34.44)
	Asian	15 (6.4)	217 (16.7)	115 (26.6)	-
	African American	10 (4.3)	67 (5.2)	40 (9.2)	78 (43.33)
	Pacific Islander	4 (1.7)	3 (0.2)	4 (0.9)	7 (3.89)
	Caucasian	192 (81.7)	419 (32.3)	243 (56.1)	-
	Other	4 (1.7)	29 (2.2)	32 (7.4)	33 (18.33)
	Latino/a/x?	No	202 (86)	1148 (88.4)	350 (80.8)
Yes	12 (5.1)	124 (9.6)	71 (16.4)	72 (46.2)	
Prefer not to answer	3 (1.3)	13 (1.0)	12 (2.8)	1 (0.6)	

<sup>a</sup>Data collection methods for “major” information varied across time point and collection location. Some instructors constrained students to 6-option forced choice selection (including “Other”), while others allowed students to list their current major as they chose. Major data were not collected in fall 2019. This accounts for some of the variation in rates of majors presented here

### Procedures

All activities conducted during this and all subsequent research steps were completed with the approval of the Emory University IRB (Protocols IRB00105275 and

IRB00114138). Subjects were recruited from STEM courses over the Spring/Summer 2018 semester. Any student enrolled in a participating STEM undergraduate course was eligible to participate. After agreeing to



**Table 4** Fit statistics for nested confirmatory factor models of the PFAI in a STEM context

Model [Result for “good fit”]	AIC [Lower]	RMSEA (90% CI) <sup>c</sup> [< 0.06]	CFI [> 0.90]	SRMR [< .08]
Conroy et al. (2002) 5 factors predicting overall FF	Not provided	0.06 (.05–.06)	0.91	0.07
Model A: Fit of Conroy’s 5 factors to undergrad students in STEM	18,802.92	0.151	0.570	0.121
Model B: Fit of modified factors <sup>a</sup> to undergrad students in STEM	15,247.87	0.150	0.638	0.112
Model C: Predicting overall FF with modified factors	15,293.97	0.153	0.614	0.120
Model D: Modified model to predict overall FF <sup>b</sup>	14,409.09	0.152	0.642	0.116
Model E: Modified model using STEM-specific items	13,965.28	0.150	0.696	0.126

Notes:<sup>a</sup> items 1, 2, 3, 9, and 10 dropped in this model; <sup>b</sup> item 12 dropped in this model; <sup>c</sup> because an imputed dataset was used, 90% CI for RMSEA were not computed by the MPlus software. Difference in fit for STEM-specific items was also tested at Models B and C. As fit did not improve significantly, values are not reported here for parsimony.

participate in the study, subjects were randomly assigned to one of two groups following a planned missingness design. That is, we intentionally provided only half of our items to 50% of the sample and the other half of the items to the other 50%. While this creates a large amount of missing data, as long as random assignment is used to decide which participants receive which half of the questions, responses for the remaining items can be imputed (Little 2013; Little and Rhemtulla 2013; Rhemtulla and Little 2012). This method of data collection was chosen to avoid survey fatigue. Students in this sample were asked to answer survey items related to the FF measure discussed in this paper and also to provide survey data on a larger number of other intrapersonal constructs of interest (i.e., coping behaviors, growth mindset). Taken together, asking students to respond to all of these items would have resulted in a survey exceeding the recommended length of 10–20 min (Cape and Phillips 2015; Revilla and Ochoa 2017). To ensure we collected high-quality data from students on all measures, a planned missingness design was judged the best approach. We also randomly assigned participants to either the “general” or “STEM” group. The “general” group received the PFAI items as they were validated by Conroy et al. (2002). The “STEM” group received versions of these same items that were preceded by language that prompted respondents to consider failures and challenges *in STEM contexts specifically*. Both groups rated the items on a scale of 1 “I believe this is never true of me” to 5 “I believe this is true of me all of the time”. We randomly assigned these versions of the survey to two groups because we wished to avoid survey fatigue by not asking students to respond to both versions and because we wished to investigate whether or not the language prompting students to explicitly

consider STEM contexts influenced FF responses. After responding to the FF questions, all students were asked to rate their level of anxiety specifically related to taking STEM courses. Finally, participants completed demographic questions. We intentionally placed the demographic questions at the end of the survey to mitigate any effects of stereotype threat that can be introduced by such questions.

## Results

### Preliminary results

**Missingness** To confirm that the intentional patterns of missingness (that is, what data are missing from participants) created as part of our planned missingness design (Little 2013; Little and Rhemtulla 2013; Rhemtulla & Little 2013) are missing completely at random (MCAR), Little’s MCAR Test (Little 1988) was computed. Results confirm that our data are missing at random ( $\chi^2(256) = 127, p > .05$ ), and it is appropriate to impute missing values. Missing values were imputed with five iterations (Schafer et al. 1997) and the imputed datasets were used for all further analyses. We calculated estimates separately for each imputed dataset and then averaged those estimates to derive final model estimates based on Rubin’s rules for multiple imputation (Rubin, 1978).

**Descriptive analyses** Outliers in the dataset were identified using the outlier labeling method (Hoaglin and Iglewicz 1987; Hoaglin et al. 1986; Tukey 1977) which labels identified outliers as “missing” to exclude them from further analysis without removing them entirely from a dataset. Visual inspection of skewness and kurtosis (George & Mallery, 2016), as well as Shapiro-Wilk’s testing indicated that our data were not normally

distributed ( $p < .05$ ; Shapiro and Wilk 1965), so the robust maximum likelihood ratio (MLR) was the most appropriate for main analyses in *MPlus* (Asparouhov and Muthén 2018; Muthén and Muthén 1998-2018). Preliminary analyses also indicated that students in our sample reported relatively low levels of FF, with dimension averages ranging from 1.46 ( $SE = .030$ ; FSE) to 3.30 ( $SE = 0.30$ ; FUF) on a 5-point scale where higher scores indicate more FF.

### Main results

Main analyses were carried out in *MPlus v. 8.1* (Muthén & Muthén, 1998-2018). Factor analyses examined the fit of a variety of nested models, beginning with a CFA of the five-factor model of FF proposed by Conroy et al. (2002). In accordance with recommendations from Knetka et al. (2019), model fit was assessed using Akaike's Information Criterion (AIC), Root Mean Square Error Approximation (RMSEA), Comparative Fit Index (CFI), and Standardized Mean Square Residual (SRMR). AIC compares each proposed model to a theoretical "true" model, calculating how far data fit to the model fall from this theoretical ideal. AIC also allows for comparison of the fit between models fit to the same sample; the AIC value for each model will be that respective model's distance from the "true" fit for the data. So, the model with the lowest AIC represents the best fit for those data (Akaike 1998; Kenny 2020). RMSEA values describe the "badness of fit," so once again a lower number is preferred. CFI assesses incremental improvements in model fit above a baseline model; thus, higher values indicate better fit (Kline 2010; Taasobshirazi and Wang 2016). And, finally, SRMR represents the standardized difference between a predicted correlation among error residuals and the actual observed correlations. Since a smaller difference between these correlation values would indicate closer convergence between prediction and observation, a smaller SRMR value indicates good fit (Kline 2010; Taasobshirazi and Wang 2016). See Knetka et al. (2019) for more complete descriptions of how each metric is calculated and their meaning. Fit statistics for all models can be found in Table 4, along with cut-off criteria used to assess goodness of fit. For all models discussed below, changes made for earlier models are carried over to later models unless otherwise stated.

**Model A: CFA of Conroy's structure** As described above, Conroy et al. (2002) proposed a five-factor structure to capture different sources of FF (Table 1). CFA was used to determine whether items in the college STEM sample loaded on the factors proposed a priori by Conroy et al. (2002; i.e. supported the proposed conceptual model). Fit statistics suggest that this model has weak to mediocre fit for students in STEM contexts (Table 4). Analysis of standardized factor loadings for individual items indicates that Question # 9 ("When I am

failing, I lose the trust of people who are important to me") does not load onto the FUIO factor ( $\lambda = 0.450$ ,  $p > .05$ ) as proposed by Conroy et al. (2002). Further investigation of beta output provided by the *MPlus* program suggests that, in our sample, the following items do not load onto *any* of the proposed factors: 1 ("When I am failing, it is often because I am not smart enough to perform successfully"; FDSE), 2 ("When I am failing, my future seems uncertain"; FUF), 3 ("When I am failing, it upsets important others"; FUIO), 9, and 10 ("When I am not succeeding, I am less valuable than when I succeed"; FSE).

**Model B: Modified factor structure** Based on the results of Model A, items 1, 2, 3, 9, and 10 were removed from the item inventory, and the CFA based on Conroy's PFAI model was rerun. Model fit improved substantially (see Table 4) though overall fit was still considered to be "poor" and individual factor loadings did not suggest that model fit would be improved by the further removal of items.

**Model C: Using modified factor structure to predict overall FF** Conroy et al. (2002) also hypothesized that their instrument would explain differences in students' overall FF. Model C tests that hypothesis using our modified factor structure (with items 1, 2, 3, 9, and 10 dropped from their respective dimensions). For this model, an additional step was added in which, after individual items predicted factor formation, the factors together predicted overall mean level of FF. We see (Table 4) that model fit worsens, but not back to the level of Model A. Also, examination of standardized beta weights suggests that the negatively coded item 12 ("When I am failing, I am not worried about it affecting my future plans") on the FUF factor is a poor fit for this sample of undergraduate STEM students when predicting overall FF ( $\beta = -0.331$ ,  $p > .05$ ).

**Model D: Modified overall model** Removing item 12 from the overall model improves model fit (Table 4) and does not yield any further suggestions for improved model fit for either the individual composition of factors or to increase the model's ability to predict overall FF.

**Model E: Modified model with STEM-specific items** Our final model in this step tested our hypothesis that question wording which primed students to *think specifically* about STEM contexts when completing the PFAI survey would lead to a better model fit. To test this, we took our best-fitting model from our work with the existing PFAI items (Model D) and substituted data from our STEM-specific questions. We assessed the effect of STEM-specific language *after* finding the best overall model fit with Conroy's original items because we wished to see if this change affected model fit above

and beyond other modifications. When these STEM-specific item variants were used, model fit improved to its highest level. While we would still not classify this as a strong model fit, it is nonetheless markedly improved and better represents the population of interest.

**Convergent validity** Convergent validity of the PFAI in general—that is, the degree to which FF measured by the PFAI is correlated with other constructs which, theoretically, should be related to FF—has been extensively addressed by Conroy and colleagues in their original validation protocol (see Conroy et al. 2001; Conroy et al. 2002). Another variable within our dataset which addresses affective components thought to be related to FF is STEM anxiety, measured via one question: “On a scale from 0 to 10, how anxious are you about your performance in your STEM classes?” Assuming that our refined model for the PFAI has good convergent validity, we would expect mean overall FF and STEM anxiety to be highly correlated. Overall, this sample reported moderate levels of STEM anxiety ( $M = 6.37$ ,  $SE = 0.175$ ). Responses ranged from 0 to 10 with a variance of 6.670, indicating that this question has sufficient variance to be used in assessments of convergent validity. Overall FF obtained by our measure is significantly correlated with STEM anxiety ( $r = 0.568$ ,  $p < .0001$ ), supporting the convergent validity of the modified PFAI.

### Brief discussion

Our initial CFA demonstrated that the PFAI best reflects university STEM students’ fear of failure when the language of the survey specifically directs them to consider their experiences within the STEM academic context. This use of STEM-specific language significantly improved model fit above the original model; however, it still did not result in a model that was well-fitting overall (Akaike 1998; Kline 2010; Taasobshirazi and Wang 2016). This implies that the underlying model structure of the PFAI might be inappropriate to assess FF in undergraduate STEM students. To explore this possibility, and to find the model structure with the greatest efficiency for measuring FF in undergraduate STEM samples, an EFA was conducted next.

### Step 2: Exploratory factor analysis (EFA) to define new model structure

In contrast to the CFA described above, EFA frees individual items from any a priori organizational constraints, allowing them to reorganize into new factors based on responses of participants, rather than researchers’ preformed hypotheses regarding how the items should cluster together. Thus, compared to CFA, which investigates whether data “fit” an existing conceptual model, EFAs suggest new models that best fit the data (Knetka et al,

2019). We hypothesized that EFA with the STEM-specific items would yield a well-fitting model of the PFAI for undergraduate STEM students by allowing removal or reorganization of some of the items among Conroy’s (2001) five proposed dimensions (described above in “Introduction”) or organization into new factors representing different dimensions. Our justification of this hypothesis is that students in STEM contexts may view failures differently than other undergraduate students. STEM professionals view failure in unique ways not generalizable to all populations (Simpson & Maltese, 2014), and STEM students describe FF as occurring as a result of specific STEM contexts and not as a more general cross-context fear (e.g., Ceyhan and Tillotson 2020; Cooper et al. 2018; Onwuegbuzie 2004), supporting the idea that FF is highly context-specific (Cacciotti 2015). Thus, the constructs proposed for other undergraduate populations may require revision for STEM undergraduate populations.

### Participants and procedures

In accordance with best practices in psychometrics, especially with regard to statistical power (Knetka et al. 2019), data for this EFA were acquired from a new dataset that included approximately 1800 undergraduate STEM students. These participants were drawn from the same research network as those in “Step 1: Confirmatory factor analyses (CFAs) of existing models,” which was expanded to include more minority serving and 2-year institutions. These data were collected in Fall 2018 as part of a pre-survey completed by students within the first month of the semester, prior to the first major assessment in their participating STEM course. The vast majority of courses included in this sample were traditionally targeted for first- or second-year students. Once the data were cleaned (e.g., outliers truncated, cases with majority missing data deleted), a sample of 1309 college students in STEM contexts remained for analyses (Hoaglin and Iglewicz 1987; Hoaglin et al. 1986, and Tukey 1977). Demographics for this sample can be viewed on Table 3, column 4. Because the STEM-specific items provided a better fit in the initial CFA study (see “Step 1: Confirmatory factor analyses (CFAs) of existing models”), students in this study were only asked to complete versions of the original twenty-five items of the PFAI which had been modified to be STEM-specific. Independent samples *t*-tests comparing the key demographics of this sample to the sample in our first analyses found no significant differences between participants on race, parents’ level of education, or reported STEM anxiety (all  $p$ ’s  $> .05$ ). There were significant differences observed between participants on age, class standing, and gender, with participants in this sample tending to be older, less academically advanced, and

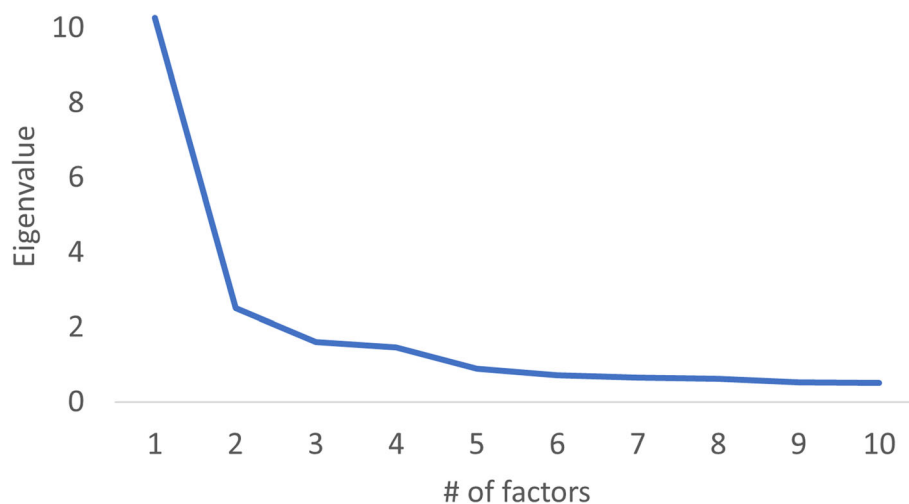
male. However, these differences were relatively small (see Table S2). Table 3 displays frequencies of other key demographic variables for all samples.

### Results

Eigenvalues and Scree plots are first steps in EFA that are used to determine how many factors a researcher should consider including in their measurement model by exploring how much variance might be explained by the addition of more factors. Eigenvalues provide a basic measure of how much unique information each assumed factor provides. For that reason, factors with higher eigenvalues are considered more useful; in general, researchers should only include factors with eigenvalues above one in their models (Knetka et al., 2019). Scree plots help provide a visual aide for this determination by plotting eigenvalues against the number of factors. Researchers should limit the number of factors at the point in the Scree plot where the curve experiences its first sharp drop (Cattell 1966; Knetka et al., 2019). These general guidelines can be widely interpreted and are meant only to help researchers limit the beginning number of factors considered for EFAs. It is important to carefully examine the quantitative fit statistics (e.g., AIC, RMSEA) generated for all potential models before making conclusions regarding the optimum number of factors or goodness of fit for any model. It is also important to consider the theory underlying the generation of survey items and the ultimate proposed use of a measurement (Knetka et al., 2019). Exploration of the Scree plot (see Fig. 2) and eigenvalues suggested that a model having between one to five factors would be the most effective for this sample of STEM undergraduates. This determination was based on established criteria of visual inspection of the Scree plot for initial leveling of

slope (Kaiser 1960) and eigenvalues greater than 1.0 (Cattell 1978). *MPlus v. 8.1* (Muthén & Muthén, 1998–2018) was used to successfully carry out EFA for each of the proposed factor structures. Model fit was assessed using Akaike’s Information Criterion (AIC), Root Mean Square Error Approximation (RMSEA), Comparative Fit Index (CFI), and Standardized Mean Square Residual (SRMR) as described in the “Results” section of “Step 1: Confirmatory factor analyses (CFAs) of existing models,” above (Kline 2010; Taasoobshirazi and Wang 2016). Model fit statistics are in Table 5.

Both the four-factor model and five-factor model provide a good fit of the PFAI items for STEM undergraduate students (Table 5). Therefore, to further investigate fit, we examined the factor structures themselves. Any item that loaded onto a factor with a loading above 0.40 and a distance of at least 0.20 from any cross-loadings was retained on that factor (Masaki 2010). Using these criteria, items that failed to load clearly onto a unique factor were dropped from the measure. From this evaluation of the factor structures, the four-factor model emerged as both conceptually and practically stronger than the five-factor model, as the five-factor model contained two factors having only one item each and a total of twelve dropped items. In contrast, the four-factor model required dropping only ten items, and the remaining fifteen PFAI items were more evenly distributed across factors that echo the original dimensions proposed by Conroy et al. (2001). Our revised form of the PFAI can be viewed in Fig. 3, and the factor loadings,  $R^2$ , and residual variances for the four-factor model are displayed in Table 6. Correlations among latent factors can be seen in Table S3 and are within acceptable bounds (Brown 2015; Watkins 2018).



**Fig. 2** Scree plot for determining number of factors in EFA

**Table 5** Model fit statistics

Model [result for "good fit"]	AIC [Lower]	RMSEA (90% CI) [< 0.06]	CFI [> 0.90]	SRMR [< 0.08]
<b>EFA for new model structure</b>				
3 Factors	88,741.92	0.080 (0.077–0.083)	0.877	0.044
4 factors	87,379.252	0.051 (0.048–0.055)	0.954	0.023
5 factors	87,162.627	0.047 (0.043–0.051)	0.965	0.019
<b>CFA of new model structure</b>				
EFA sample (Fall 2018)	52,459.976	0.054 (0.049–0.060)	0.960	0.040
Second CFA sample (Fall 2019)	16,733.419	0.060 (.051–0.070)	0.964	0.049
Initial CFA sample (Summer 2018)	52,459.98	0.054 (0.049–0.060)	0.960	0.040
PEER sample	5971.960	0.071 (0.051–0.090)	0.936	0.058

**Brief discussion**

Use of EFAs allowed us to further refine the PFAI for use in college-aged STEM student samples. Our best-fitting model significantly reduced the number of items from twenty-five to fifteen, which may aid with increasing compliance and decreasing cognitive load when surveying college-aged STEM students (reviewed in Peytchev and Peytcheva 2017). Based on a series of qualitative cognitive interviews with STEM undergraduates (see *Step 7*, below), it appears that several of the dropped items contained

words or phrases that made them unclear or ambiguous to students. One dimension, FDSE, contained items with the words "talent," "hate," and "not in control" to describe situations in which students might devalue their own self-estimate. In cognitive interviews, students objected to these words, and ultimately, FDSE was not supported by factor analysis as a unique dimension. It is likely the wording of items did not align with STEM students' views of themselves when responding to failures. Interestingly, one item was retained in the scale from the original group of

I believe this is never true of me. (1)	I believe this is sometimes true of me. (2)	I believe this is true of me half of the time. (3)	I believe this is true of me most of the time. (4)	I believe this is true of me all of the time. (5)
<i>Fear of an Uncertain Future</i>				
2. When I am failing, my future seems uncertain.				
4. When I am failing, I blame my lack of talent.*				
5. When I am failing, I believe that my future plans will change.				
8. When I am failing, it upsets my "plan" for the future.				
<i>Fear of Upsetting Others' Losing Interest</i>				
11. When I am not succeeding, people are less interested in me.				
13. When I am not succeeding, people seem to want to help me less.				
17. When I am not succeeding, people tend to leave me alone.				
21. When I am not succeeding, some people are not interested in me anymore.				
23. When I am not succeeding, my value decreases for some people.				
<i>Fear of Upsetting Important Others</i>				
3. When I am failing, it upsets important others.				
14. When I am failing, important others are not happy.				
19. When I am failing, important others are disappointed.				
<i>Fear of Experiencing Shame and/or Embarrassment</i>				
18. When I am failing, it is embarrassing if others are there to see it.				
24. When I am failing, I worry about what others think about me.				
25. When I am failing, I worry that others may think I am not trying.				

Note: Item numbers refer to original placement in Conroy et al.'s (2002) brief PFAI

\* Item originally loaded on *Fear of Devaluing One's Self-Estimate* sub-scale

**Fig. 3** Modified version of the PFAI

**Table 6** Factor loadings for four-factor model of PFAI

Item	Factor loadings <sup>a</sup>				R <sup>2</sup>	Estimated residual variance
	FUF	FIOLI	FUIO	FSE		
When I am failing, my future seems uncertain.	<b>.795</b>	.393	.313	.468	0.74	0.380
When I am failing, I blame my lack of talent. <sup>b</sup>	<b>.714</b>	.424	.307	<b>.502</b>	0.57	0.533
When I am failing, I believe that my future plans will change.	<b>.814</b>	.366	.343	.417	0.77	0.298
When I am failing, it upsets my “plan” for the future.	<b>.804</b>	.368	.409	.486	0.76	0.320
When I am not successful, people are less interested in me.	.413	<b>.807</b>	.461	.489	0.72	0.322
When I am not successful, people seem to want to help me less.	.301	<b>.733</b>	.406	.292	0.70	0.493
When I am not successful, people tend to leave me alone.	.323	<b>.736</b>	.411	.365	0.71	0.461
When I am not successful, some people are not interested in me anymore.	.361	<b>.846</b>	.492	.424	0.84	0.268
When I am not successful, my value decreases for some people.	.411	<b>.826</b>	.462	.476	0.74	0.303
When I am failing, it upsets important others.	.280	.276	<b>.689</b>	.260	0.81	0.563
When I am failing, important others are not happy.	.273	.471	<b>.820</b>	.329	0.92	0.332
When I am failing, important others are disappointed.	.304	<b>.515</b>	<b>.875</b>	.413	0.78	0.203
When I am failing, it is embarrassing if others are there to see it.	.446	.412	.439	<b>.778</b>	0.75	0.364
When I am failing, I worry what others think about me.	.489	.446	.378	<b>.855</b>	0.84	0.242
When I am failing, I worry that others may think I am not trying.	.417	.348	.379	<b>.698</b>	0.50	0.514
<b>Dropped items<sup>c</sup></b>						
When I am failing, it is often because I am not smart enough to perform successfully.	<b>.510</b>	.263	.206	.293		
When I am failing, I expect to be criticized by important others.	.416	.450	<b>.618</b>	.498		
When I am failing, I am afraid that I might not have enough talent.	<b>.752</b>	.386	.265	<b>.588</b>		
When I am failing, I lose the trust of people who are important to me.	.380	<b>.655</b>	.521	.227		
When I am not successful, I am less valuable than when I succeed.	<b>.554</b>	.632	.413	<b>.550</b>		
When I am failing, I am not worried about it affecting my future plans. <sup>d</sup>	–	.178	–	–		
	.254		.029	.149		
When I am not successful, I get down on myself easily.	<b>.598</b>	.291	.290	<b>.691</b>		
When I am failing, I hate the fact that I am not in control of the outcome.	<b>.512</b>	.291	–	<b>.599</b>		
			.233			
When I am failing, I believe that everybody knows I am failing.	.470	<b>.533</b>	.403	<b>.720</b>		
When I am failing, I believe that my doubters feel they were right about me.	.462	<b>.622</b>	.411	<b>.675</b>		
<b>Cronbach’s alpha</b>	.832	.854	.818	.730		
<b>R squared</b>	.482	.606	.481	.625		

<sup>a</sup>Factor loadings over .50 appear in bold. Factor loadings below .30 appear in lighter gray font

<sup>b</sup>This item original loaded on the FDSE subscale; all other items load onto their original subscale

<sup>c</sup>Items were dropped from the scale for one of two reasons: (1) no factor loadings were above .40 or (2) the difference among factor loadings above .40 was less than .2, indicating too high a degree of cross-loading

<sup>d</sup>This item is reverse scored (i.e., higher scores actually indicate lower FF, contrary to all other items on the scale)

questions for the FDSE dimension. “When I am failing, I blame my lack of talent,” now loads onto the FUF factor. This suggests that perhaps STEM students at the college level view talent as a potential advantage (or stumbling block) for future success, rather than a reflection on their current self-estimate.

Ultimately, four of the five Conroy-proposed dimensions were still represented in the final fifteen-item revised model, and a majority of items present loaded onto their “original” dimension (Fig. 3), with four items

assessing Fear of an Uncertain Future (FUF), five items assessing Fear of Important Others Losing Interest (FIOLI), three items assessing Fear of Upsetting Important Others (FUIO), and three items assessing Fear of Shame and Embarrassment (FSE). It is worth noting that several residual variances for the individual items in the model remain high (Table 6). This suggests that there is still some variability in students’ responses to the PFAI items that is not explained by the current model. Thus, it is possible that more factors could remain to be

extracted (Pett et al., 2003; Watkins 2018). More qualitative work within a STEM context could elucidate additional FF dimensions relevant for STEM undergraduates which could augment the current instrument in future iterations of scale development.

### Steps 3–5: CFAs to confirm fit of new factor structure

After determining a new factor structure, it was important to verify that the structure fit well in more than only the sample of STEM students used to conduct the EFA. We subsequently performed a series of CFAs using similar methods to those described above on this newly suggested structure.

#### Step 3

We first verified the fit of our model within the sample of students used for the EFA in *Step 2*. This sample included 1309 students recruited from the FLAMENet research network during Fall 2018 (Table 3, column 4). CFA within this sample yielded excellent model fit (Table 5, row 7).

#### Step 4

We next wanted to confirm the revised factor structure in a separate sample of STEM undergraduates to verify the stability of the model. FLAMENet participants during Fall 2019 provided data on 433 students (see Table 3, column 5 for demographics). This analysis also proved to have excellent fit (Table 5, row 8).

#### Step 5

Finally, we wanted to return to our original sample of 235 students recruited during Summer of 2018 (Table 3, column 3) to see if the revised model provided good fit given that our efforts with CFA in “Step 1: Confirmatory factor analyses (CFAs) of existing models” improved model fit significantly but did not reach the threshold of good fit. The model demonstrated excellent fit in this sample as well (Table 5, row 9).

### Brief discussion

We conducted three separate CFAs to verify that the new model structure for the PFAI indicated by the EFA (*Step 2*) could be replicated in multiple samples of STEM undergraduates. In all three cases, fit statistics indicated excellent model fit. Independent samples t-tests among the various samples identified some significant differences among demographic variables (see *Table S2*). These differences, while statistically significant, were small, and the model’s continual good fit despite them demonstrates its robustness as an assessment tool across undergraduate STEM samples.

### Step 6: Model fit among persons excluded because of their ethnicity or race (PEERs)

The work described thus far represents a novel presentation of the PFAI which we have shown to be a stronger fit for undergraduates’ actual conceptualization of FF in STEM contexts. While work aimed at improving the validity and applicability of interventions and assessment represents a worthy goal of service for all students, it is especially salient for PEER students, who are more likely to leave STEM academic programs (Asai 2020; National Science Board 2018; Steele 1997; Stinebrickner and Stinebrickner 2014). Factors such as FF are likely to be important leverage points for improving STEM students’ ability to persevere through academic challenges and failures. The implied long-term impact of aligning pedagogical practices to reduce FF is increased inclusion and success in STEM education and careers (e.g., retention within STEM majors, Nelson et al. 2019). For the PFAI to be an effective assessment tool, then, it is critically important to ensure that the same factor structure is valid for people at higher risk of leaving STEM, such as PEERs (Asai 2020). Additionally, previous intervention studies with psychological constructs that influence students’ responses to challenge and failure (e.g., mindset) suggest that these interventions may be most effective for PEER students (Aronson et al. 2002; Fink et al. 2018; Yeager et al. 2016). Thus, we conducted separate model fit analyses with a sample of only PEER undergraduate STEM students to explore how this instrument functions when assessing this critically important population.

### Participants and procedures

Participants for this analysis were drawn from the same dataset of approximately 1309 undergraduate STEM students described in the previous EFA section (*Step 2*), along with two other datasets collected in Fall 2019 (*Step 4*) and Summer 2018 (*Step 1*). While these data were pulled from surveys collected at different times, there were no differences in the method by which surveys were presented. These data were then coded to classify each student as either a “PEER” (1) or “not a PEER” (0). Any student who self-identified as “White/Caucasian” or “Asian” on a demographic survey question was not considered a PEER; all other students were coded as a PEER. This classification was based on data from the NSF which indicates that Asian students are not typically underrepresented in STEM and health-related sciences in the USA (Asai 2020; National Science Foundation 2020). In total, 280 PEER students were identified. Full demographics for this sample are included in Table 3, column 6. In our sample, PEER students identified as belonging to African American or Black; American Indian or Alaskan Native; Arabic or Middle Eastern; Hispanic or Latinx; and/or some other racial or ethnic group. All

further analyses described in this section have been conducted with *only those* students classified as PEERs. Independent samples t-tests (see Table S2) verify that this combined sample of PEER students contained a significantly higher number of PEER students than the samples from which it was drawn (by an average factor of 10). Small differences existed among other demographics, but in general PEER demographics were intermediate or roughly equivalent to other samples (Table S2) with all differences being small. The only difference which might also have impacted students' FF was that PEER students reported that their parents received a lower overall level of education compared to students in our other samples. This has often been used as a proxy for socioeconomic status (SES; Pascarella and Terenzini 1991; Snibbe and Markus 2005) and could mean PEER students are under financial stress, making them more likely to fear failure. However, PEERs, in general, are more likely to hail from first-generation or lower SES backgrounds in the USA (Cullinane 2009; Kuh et al. 2006); thus, this observed difference is not surprising. Indeed, differences in educational background that correlate with race and ethnicity, in part, contribute to the need to understand, study, and create measures specific to PEER groups. PEER students also reported equivalent levels of STEM anxiety, which suggests they are not, overall, more anxious about STEM courses than non-PEER STEM students.

## Results

*MPlus v. 8.1* (Muthén & Muthén, 1998-2018) was used to conduct CFA for the four-factor, 15-item model described above (see Tables 4 and 5). Model fit was assessed using AIC, RMSEA, CFI, and SRMR as described in the "Results" section of "Step 1: Confirmatory factor analyses (CFAs) of existing models" above (Kline 2010; Taasobshirazi and Wang 2016). Model fit statistics are displayed in Table 5, row 10. All fit statistics are within acceptable ranges for a "good" fitting model (Kline 2010; Taasobshirazi and Wang 2016). While the RMSEA value of 0.071 slightly exceeds our established criterion of  $RMSEA < 0.06$  for a "good" fitting model, the 90% confidence interval does include this value. In addition, disagreement abounds regarding appropriate cut-off points for fit indices (e.g., Hayduk et al., 2007), with some researchers arguing that RMSEA can rise as high as 0.08 before a model is considered a "poor" fit (MacCallum et al. 1996). RMSEA, along with SRMR, is also the fit index perhaps most susceptible to inflation with small sample sizes (Kenny et al. 2015). The difference in sample size between samples in *Step 2* ( $N = 1309$  in the full EFA) and *Step 6* ( $N = 280$  in our CFA with PEER students) may partially explain the increase in RMSEA.

## Brief discussion

By conducting this sub-analysis, we demonstrate that our 15-item modified form of the PFAI does provide a statistically good fit for PEER students in STEM. Given past research on both the disproportionate difficulties of pursuing a STEM career as a PEER student and the increased effectiveness of interventions for PEER students (e.g., Sisk et al. 2018), this implies that the modified PFAI could be an especially powerful assessment tool for future research. More research is needed to assess if this is a broad effect across all classes of underrepresented and excluded identities. Our sample was restricted to racial and ethnic exclusion and, even then, all possible identities were not represented (e.g., we did not collect data on subgroups within the broad category "Asian"). In addition, other types of underrepresentation and exclusion, such as gender, sexual orientation, first-generation status, and religious affiliation, likely influence FF and may affect responses on the PFAI. Future studies should investigate the fit of our modified PFAI among these groups and for students with identities that intersect multiple underrepresented groups.

## Step 7: Cognitive interviews

Cognitive interviews were conducted to assess face validity of all twenty-five items proposed by Conroy et al. (2001). Face validity describes the extent to which a test or survey measures what it purports to measure. Cognitive interviews are an excellent way to assess face validity of survey questions measuring latent intrapersonal constructs because they allow the researcher to directly ask participants about their interpretation of survey items and to then assess whether this interpretation matches with the intended purpose of the item. They also assess participants' understanding of the content of the instrument in addition to elucidating what the participant is thinking and feeling while responding, which can often influence the valence of responses (Willis 2015). For our study, we used cognitive interviews as the last step in our data collection to (a) check the face validity of our items (were the items interpreted by STEM undergraduates as we intended), (b) help elucidate potential reasons that certain items did not have good fit in our CFA and EFA analyses, and (c) provide clarity and additional information about how students were interpreting certain phrases that were more ambiguous in the PFAI items.

## Participants and procedures

In accordance with results from our initial CFA (Step 1), students who participated in cognitive interviews were asked to consider the wording of the PFAI questions specifically in the context of their STEM courses and research (e.g., "For the following



questions, please consider challenges and failures that you face specifically in your STEM course(s) and research”). Eleven students completed interviews of approximately 20 min each via Zoom in return for a \$20 Amazon gift card. The research opportunity was announced to students by FLAMENet instructor partners. Interested students completed an initial screening questionnaire. From this information, the research team selected students to participate in a deliberate attempt to achieve a sample with approximately equal gender and racial distribution and who represented STEM fields similar to those seen in our larger sample(s). Demographic characteristics of these students are described in Table 7. During these interviews, students were asked (a) if the meaning of each question was clear and how they interpreted the question, (b) if answer choices seemed appropriate, (c) if there were any suggestions for improving the question, and (d) if they had any other thoughts. In addition, students were asked to clarify their thought process related to the somewhat ambiguous phrase “important others,” which appears in many of the items (e.g., “Which specific people come to mind when they hear this phrase?” “Is it the same or different for each question which uses this phrase?” etc.). Students were also asked to provide their thoughts on three questions added by the researchers prior to the PFAI items on the survey. These items asked respondents

to describe a time when they recently encountered a challenge or failure in their STEM course(s) and then to rate how upsetting they found that event on a scale from 0 to 10, with 10 being the most upsetting. Students were also asked to report how anxious they were about their performance in their STEM course(s) on a scale from 0 to 10, with 10 indicating the highest levels of anxiety (see “STEM anxiety” under “Step 1: Confirmatory factor analyses (CFAs) of existing models”, above).

### Results

Overall, students found the survey instructions clear and appropriate. Students reported that our additional questions, which asked them to describe a recent challenge or failure in a STEM context and to rate how upsetting that experience was, along with their general levels of anxiety in STEM contexts, did not prompt any confusion, discomfort, or concern during the interviews.

Student responses to the ten items dropped through EFA (Step 2) both support the removal of these items from the measure and provide some explanations for why these items may not fit well for students in undergraduate STEM contexts. With a majority of these items, there appear to be specific words or phrases that generate confusion for the respondent. For example, with item 7, “When I am

**Table 7** Demographic characteristics of participants in cognitive interviews

Student <sup>1</sup>	Gender	Age	Class standing	Major	Race	STEM anxiety (0–10)
CE	Male	21–23	4th Year	Chemistry	White or Caucasian	4
DQ	Male	18–20	4th Year	Biology	White or Caucasian	1
HS	Female	18–20	3rd Year	Integrative Physiology; Language Sciences	White or Caucasian	3
HA	Male	18–20	2nd Year	Chemistry; Psychology	Asian	4
LF	Female	21–23	4th Year	Nuclear Medicine	American Indian or Native Hawaiian; Black or African American	10
MZ	Female	21–23	4th Year	Engineering	Black or African American	6
MB	Female	21–23	4th Year	Psychology	Asian	9
PO	Male	18–20	3rd Year	Biology	American Indian or Native Hawaiian; Black or African American	8
PW	Male	18–20	3rd Year	Chemistry	Black or African American	9
UN	Female	18–20	3rd Year	Chemistry	White or Caucasian	6
VH	Female	21–23	4th Year	Engineering Management	Black or African American	4

<sup>1</sup>Student pseudonym initials follow quotes from cognitive interviews

failing, I am afraid that I might not have enough talent,” students expressed hesitation with the word “talent,” especially since they felt it did not describe STEM contexts.

Why “talent”? I would expect to see “intelligence.” I’m not used to thinking about talent in this context [science and STEM].—CE

Talent—is that specific for that subject? It’s not as clear as “smart enough”? Talent is associated more with singing, dancing, etc.—MB

Similarly, for item 16, “When I am failing, I hate the fact that I am not in control of the outcome,” students objected to the use of “hate,” often stating that hate was too “strong” a word. They also felt that “not having control” was inappropriate in this context.

My initial reaction is that you’re always in control a little bit; I just don’t think anyone is not in control of the outcome. Is there a different way to word this?—CE

For the other dropped items, students expressed similar objections to specific words or phrases they found confusing.

General themes in students’ interview responses also provide important insight into how students interpreted the survey items. Throughout the measure, the item stems “When I am failing” and “When I am not succeeding” are alternated and, presumably, are thought to be synonymous. However, students said that they interpret these two phrases differently and would actually have responded to some questions differently, had the opposite stem been used.

“Not succeeding” is more broad than “failing.” Failing is an “F” vs. not succeeding is not getting straight 100s when [you] wanted to. Depending on what your standard was for “succeeding,” it might change [your] response.—UN

One student expressed that these changing stems were useful because they could respond about a broader range of experiences instead of only responding about the more extreme scenario of failing, which they narrowly defined as getting an “F.”

I like the changing stems because ‘failing’ and ‘not succeeding’ are two different things. You can ‘not succeed’ without ‘failing’. You could just be doing not as good as you thought you could do. [Getting a] ‘B’ instead of an

‘A’. [You’re] not failing though, because it’s not an ‘F’. I like that both are assessed with these questions.—PO

This conflict can especially be seen in student responses to item 10, the first time that the phrase “not succeeding” is introduced as an alternative question stem to “failing” in the original Conroy structure: “When I am not succeeding, I am less valuable than when I succeed.”

Wording change to ‘not succeeding’ is weird. I had to read it twice.—MB

This change could explain why this item was one of the ten items dropped by EFA. Multiple students expressed confusion at this word change and a need to reread the question. However, as indicated by the above students’ responses, students were generally able to interpret this phrase after considering it and used it to broaden the scope of scenarios they responded about.

Finally, these interviews help provide insight into the identities and roles of individuals the students called to mind when they encountered the phrase “important others.” Interestingly, student responses suggest that the specific people brought to mind by this phrase may change depending on the actions being attributed to those important others. When important others were described as losing interest (e.g., items 11 or 21), students described current or future professors, research supervisors, and employers (“they may ‘give up’ on you”—CE) or friends and classmates (“maybe [they] wouldn’t want to study with you anymore”—LF). Describing important others as upset prompted students to think more broadly, with answers including professors, family, and friends. However, the language of some of these items specifically pointed students towards family. For example, item # 3, “When I am failing, it upsets important others,” elicited the wide range of responses previously mentioned. However, item # 19, “When I am failing, important others are *disappointed*,” keyed students into thinking more specifically of “mostly family and relatives/caretakers” (PW). Similarly, item 6, “When I am failing, I expect to be criticized by important others,” was largely associated specifically with instructors and others with academic authority such as “teachers, professors, and mentors” (CE) and also “academic advisors, tutors, etc.” (MZ). In general, it appears that the most salient “important others” are brought to mind for various aspects of the STEM academic context with these items. That is, students tend to think of the important others that are most likely, in their estimation, to experience a given emotion or respond in a specific way to their failures or lack of success.

### Brief discussion

Overall, results from our cognitive interviews support the general structure of the revised survey. There were no major confusions or issues with instructions or overall question wordings. Student responses supported the statistical decision to remove dropped items. In exploring student responses to items involving “important others,” we found that students may think of different people depending on the particular aspects of the STEM context that are evoked by the action phrases of the item (e.g., “criticize” vs. “upset” vs. “disappoint”). This ability of items to draw on the most salient people in students’ minds grants flexibility to these questions and reduces concerns that the phrasing of “important others” might be restrictive or otherwise confusing for respondents. This phrasing allows students to consider a broad range of relationships and histories among an individual student and those they consider to be “important.” However, it restricts survey interpretation in some ways because we cannot know the exact identity of the “important other” that comes to mind for students. We can only assume, based on these results, that that important other is an important person that is also likely to be perceived by the student as responding in accordance with the question language (e.g., being disappointed, upset, critical). Finally, these interviews suggest that some students do not view the phrases “when I am failing” and “when I am not succeeding” as interchangeable. However, this may, in fact, be a benefit of the measure. In most cases, students view “failing” as more negative, damaging, and permanent than “not succeeding.” By using the more mild “not succeeding” items, this measure may allow one to assess FF (or fear of not performing to a specific standard) in students who have rigid definitions for what constitutes “failure.”

### Limitations

As with any research aimed at instrument validation, this work has several limitations. A priori power analysis using *GPower 3.0* (Erdfelder et al. 1996) indicated that a sample size of 500 would be ideal for our initial planned CFA. While we recruited close to 500 participants in Step 1 ( $N = 423$ ), only 54% of these participants ( $N = 235$ ) provided data that were complete enough for analysis. This sample size limits our power to detect the small yet meaningful differences (Little 2013), which are increasingly recognized as large effects in the educational community (Kraft 2020). This may have affected model fit in our initial CFA, as fit indices are influenced by sample size (The precise effect varies by fit index; Kyriazos 2018). However, since our total sample for the initial CFA still exceeded the level at which the most conservative fit statistics begin to be affected ( $n = 200$ ) and our CFA model had many indicators estimating

each factor (5 items per factor in the original model), it is unlikely that our sample size influenced fit statistics to such an extent that erroneous conclusions were drawn (Boomsma and Hoogland 2001; Marsh and Hau 1999). In addition, our knowledge of likely recruitment difficulties led to our choice to use a planned missingness design (Little and Rhemtulla 2013; Rhemtulla & Little 2013), which resulted in the imputation of large sections of our data. This limitation is not present in our EFAs, which had a much larger sample size of 1309 and involved no data imputation. Likewise, our samples for the CFAs of our modified factor structure in a novel mixed sample and PEER-only sample did not involve multiple imputation or planned missingness. However, they both fell short of 500 participants ( $N = 433$  and  $N = 280$ , respectively).

All of our mixed samples (used for *Steps 1–5* and *7*) contained low levels of academic, racial, and gender diversity. In particular, our samples contain a majority of students identifying as female. While women do currently comprise approximately half of the undergraduate STEM population and these percentages are higher in the life sciences (National Science Board 2018), female students are still likely overrepresented in our sample. While we were able to conduct a separate fit analysis for PEER students, racial and ethnic diversity of the other samples overall were not completely representative of national trends (U.S. Department of Education, 2012) and there may be finer grained variations between students from different racial and ethnic groups that our data are not able to elucidate. In addition, in our PEER analysis Asian students were treated as non-PEERs based on NSF’s definition of Asian as not underrepresented in STEM. While this is true for the broad category, it does not take into account different Asian subgroups (e.g., Korean, Vietnamese) which may be underserved and excluded in STEM. If the goal is to assess interventions which would target underserved populations in STEM, it is especially important that assessment measures, such as the PFAI, accurately assess members of *all* underserved populations. Our investigation of the modified PFAI fit for PEERs starts this, but only scratches the surface. Future studies of the utility of the modified PFAI should consider nuances among PEER groups and other types of underserved groups (e.g., first-generation students). In addition, our sample contained a majority of Biology and Chemistry students and did not represent as many students from other STEM disciplines (e.g., Physics, Geoscience, Computer Science, Psychology). This should be taken into account when interpreting the results of this work. Also, a significant majority of participants across all samples reported pursuing a STEM major (e.g., Biology, Chemistry, Engineering). While the language modifying the PFAI is not specific to STEM

majors and simply asks respondents to consider their responses to failure and challenges within *STEM contexts*, which may be equally applicable to STEM majors and students pursuing other majors who enroll in STEM classes to fulfill graduation requirements, it is nonetheless possible that the reason for enrolling in the course influences students' goals within the STEM context and affects their responses to the PFAI. Future work should more carefully consider this possibility.

Finally, none of these samples were randomly selected. Participation was voluntary, with announcements of the research opportunity disseminated by instructors who value good pedagogy and may have previously encouraged more adaptive outcomes (like lower FF) in their classrooms. This, combined with possible self-selection of the most motivated or achievement-driven students among these classrooms, may have biased our participant group. However, our concerted efforts to collect data across multiple disciplines and multiple institutions representing diverse contexts may have partially mitigated this selection bias.

## Discussion

The aim of this study was to evaluate, revise, and present a modified version of an existing instrument, the PFAI, for STEM undergraduate populations. This work is particularly important since prior evidence suggests that FF may contribute to STEM student procrastination (Onwuegbuzie, 2004; Zhang et al., 2018), threaten motivation (Ceyhan and Tillotson 2020), and even lead to attrition from STEM (Nelson et al., 2019). Our results support the modification of the original version of the PFAI to effectively measure STEM-specific FF. Our analyses supported a more parsimonious reduced scale: fifteen items as opposed to twenty-five items and four factors corresponding to different dimensions as opposed to five. In addition, we found support for the hypothesis that STEM-specific items provide the best fit for STEM undergraduate students. Notably, our reduced, STEM-specific scale functions well for both PEER and non-PEER samples and has good face validity. We also present evidence that our reduced, STEM-specific scale estimates different levels of FF in STEM student samples than the original Conroy measure—a finding supporting the importance of the scale's revision and modification. Our results can be used to guide the use and interpretation of our new STEM-specific version of the PFAI within STEM undergraduate contexts.

### A reduced, more parsimonious, version of the PFAI

Shorter, more parsimonious scales are generally preferred as they help to mitigate survey fatigue, improve response rates, and increase measure accuracy (reviewed in Peytchev and Peytcheva 2017). Based on our factor

analyses, we were able to drop ten of Conroy's original twenty-five items, including one entire dimension (FSDE) from the measure, resulting in a shorter, more parsimonious, scale of fifteen items. Our final best-fitting model of the PFAI specifically includes the FUF, FIOLI, FUIO, and FSE scales. We assert that Conroy's (2001) original definitions of these dimensions continue to be appropriate for use in STEM populations since our cognitive interviews indicated that the items retained in the scale had reasonable face validity.

After our final examination of the EFA, the FDSE dimension proposed by Conroy et al. (2001) did not emerge among the responses of STEM undergraduates. This may mean that, within STEM academic contexts, undergraduates are not worried about damage to their self-estimate as a result of failure. However, this is not well supported in the literature, which implies that students who identify with a field of study may experience lower self-efficacy as a result of STEM failures (Bandura et al., 1999; Pajares 2005). Alternatively, it could suggest that the current PFAI items for this dimension (e.g., "When I am failing, I am afraid that I may not have enough talent.") do not accurately articulate threats to students' self-estimate within STEM contexts. Indeed, responses to cognitive interviews support this latter view, as students expressed confusion over considering "talent" in regard to STEM, as opposed to a more arts-based environment. Several studies provide evidence that people have different views of whether "talent" or similar attributes such as "brilliance" are determinants of success in certain fields (Leslie et al. 2015; Storage et al. 2016). It could be that, for the STEM fields included in this study, "talent" is not seen as a determinant of failure or success, and therefore, it did not make sense when included in some items. However, it is of note that one item that was originally on the FDSE factor ("When I am failing, I blame my lack of talent") loaded onto the FUF factor based on STEM students' responses. So, while some students were uncertain if "talent" was the right word to use when discussing science, their responses to the survey nevertheless indicate that feeling as if they possess a lack of talent in STEM contexts is linked to future uncertainty. Additional studies could address whether STEM students do, or do not, experience fear of devaluing their self-estimate when experiencing STEM failures and how students might describe this using words other than "talent" in order to generate potential new items within this dimension.

Interestingly, while all of Conroy's other proposed dimensions still emerged from students' responses, we noted that items dropped from the scale often included items with particularly strong and direct wording (e.g., "When I am failing, I *expect to be criticized* by important others" or "When I am failing, I *believe that everybody*

knows I am failing). Responses to such questions may have been impacted by individuals' tendency to, knowingly or unknowingly, underreport thoughts, feelings, or behaviors which run contrary to established social norms or are perceived to invade their privacy (Gnambs and Kaspar 2014). For example, Fisher (1993) showed that social desirability bias in survey responses was affected by direct versus indirect questioning. Indirect questions, which ask subjects to respond from the perspective of another person, alleviated social desirability bias while direct questions did not. In addition, research has found that bias tends to be enhanced when respondents view survey questions as sensitive or seeking to invade their privacy (Gnambs and Kaspar 2014; Krumpal 2013).

While the PFAI questions are not indirect, it could be that the most strongly worded PFAI questions do not load well onto factors because they more blatantly confront the respondent with constructs which are viewed as too personal or extreme to endorse. In addition, if respondents associate FF constructs with social norms and a related potential to generate personal discomfort or negative reactions, they could be less likely to endorse these beliefs (Fisher, 1993). Our cognitive interviews support these conclusions as students were often opposed to strongly valenced words such as "hate," words that carried specific connotations such as "talent," and phrases that brought into question their personal agency or privacy such as "not in control" and "everybody knows." Given our findings from cognitive interviews and EFA analyses, in addition to findings from other studies, we feel that removal of the ten items improves the modified PFAI scale not only because it makes it shorter, but also because it may avoid introducing biases as a result of emotional reactions to question wording.

#### **A STEM-specific version of the PFAI**

This work demonstrates support for our hypothesis that a STEM-specific version of the PFAI is more appropriate to measure FF in undergraduate STEM students than a non-STEM-specific version. Conroy and colleagues' original factor structure model for the PFAI did not yield a good fit with the undergraduate STEM student sample. Model fit improved significantly by removing explicit items and the one reverse-code item and by adding language to the items which specifically evoked the STEM-specific academic achievement context. This version of the PFAI corresponded better to data provided by undergraduates enrolled in STEM than the original PFAI model (Conroy et al. 2002; AIC of 13965.28 vs. 18802.92). However, even our best-fitting model from the first round of CFA analysis (Step 1) was not what is considered to be a "good" fitting model (Akaike 1998; Kline 2010; Taasobshirazi and Wang 2016). Thus, we

moved forward by allowing the PFAI items to reconverge into a new factor structure via an EFA (see Step 2) which yielded a better fitting model for STEM undergraduates (Steps 2–6). Our final model, which uses STEM-specific language to introduce the items, provides a "good" fit and can be used reliably to measure the listed dimensions of FF (Akaike 1998; Kline 2010; Taasobshirazi and Wang 2016).

Given that FF is specific to defined achievement contexts, varies from context to context (Cacciotti 2015; Conroy 2001), and is seen in nuanced ways by STEM professionals (Simpson and Maltese 2017), it is not surprising that prompting students to consider the PFAI items *within a STEM context* led to a better fit than considering the items in a non-specific context. It is also not surprising that the organization of the PFAI items into each dimension needed revision when used with a STEM audience. This context specificity is not unique to FF. Other intrapersonal elements such as mindset (Dweck 2006) and ability to cope (Lazarus 1993; Skinner et al. 2003) are known to be context specific. Yet, much like FF, scales that measure such contexts are often written for broad contexts and general audiences (e.g., Carver 1997; Dweck 2006). Given improvements to model fit for FF when a STEM context was considered, it is worth asking whether other measures of intrapersonal elements also need to be refined to address more specific contexts and how specific those contexts need to be for measurement accuracy. We also feel that there is continued need for work on how to best assess FF in STEM populations. The STEM-specific version of the PFAI that we present as a result of this work is useful for measuring the proposed dimensions. However, additional qualitative interview studies could help elucidate whether there are additional dimensions of FF unique to STEM students that should be added to more completely measure the construct as a whole (Knekta et al. 2019). This, however, is beyond the scope of the current work.

#### **The STEM-specific version of the PFAI: valid for use with undergraduate PEER populations**

Results of our PEER CFA analyses (Table 5, row 10) revealed that the modified PFAI was a good fit for a sample of solely PEER students. Likewise, our cognitive interviews, which included PEER and non-PEER students, did not uncover differences in interpretations of items that could be attributed to race or ethnicity. Thus, we assert that our modified STEM-specific version of the PFAI can be used effectively to measure FF for STEM PEERs. We consider this result a highly important finding of our work. As outlined in our introduction, PEERs have historically been excluded from STEM fields, and they continue to experience difficulties and

leave (or be excluded from) STEM at higher rates than their majority peers (Asai 2020; Huang et al., 2000; NCSSES, 2019). PEER leaving is highly detrimental to STEM and society since losing diversity threatens scientific creativity and innovation (Freeman and Huang 2014; Page, 2007), and may result in scientific communities failing to address questions that are relevant and important for PEER communities (Hacker 2013). Notably, PEER departure from STEM continues to happen despite decades of research and efforts to better understand what helps PEER students to persist (e.g., Barbosa 1975; Chang et al. 2014; Estrada et al. 2011; Estrada et al. 2016; Estrada et al. 2018; Estrada and Matsui 2019; Hurtado et al. 2010; Maton et al., 2009; Matthews 1990) and consistent support for programs that assist PEER students (e.g., HHMI Inclusive Excellence, NSF INCLUDES). And, attrition of PEER students occurs at disproportionately higher rates from STEM majors than other fields of postsecondary study (Riegler-Crumb et al., 2018). It is clear that we do not understand the whole picture. However, we do know that STEM failures or lack of achievement to an expected standard may precipitate a decision-making process that results in PEERs leaving STEM or disengaging from STEM challenges (Corwin et al. 2020; Henry et al. 2019). In addition, FF may contribute to that critical decision to leave STEM (Nelson et al. 2019). FF, then, may be a key component in understanding how “failing” (or even failing to meet one’s own high expectations) can affect PEER students’ success and persistence.

To understand how this construct acts, it is crucial to be able to accurately measure it for PEER populations, and we cannot assume that an instrument valid for the majority of undergraduates is valid for PEERs. As Knekta et al. (2019) explain, “validity is not a property of the measurement instrument,” rather it describes how well that instrument functions for a specific population in a specific context (p. 2). Any conclusions drawn from the instrument are only as strong as the instrument itself (Cronbach and Meehl 1955). Therefore, the evidence that the modified STEM-specific PFAI functions well for PEER students is an important result of this work and can be applied in future research on developing inclusive practices in STEM education. However, we wish to again highlight that more work can be done to further understand FF in PEERs. Extensive interview studies with PEERs specifically may be able to uncover other dimensions of FF specific to PEERs. This, unfortunately, is beyond the scope of our work. If this work were to be done, however, additional dimensions specifically relevant to PEERs could be added to our version of the PFAI to further improve STEM PEER FF measurement.

### Differences in estimated FF among models

When embarking on this work, we predicted that FF, and specifically the constructs measured by the items in the PFAI, might be interpreted differently by STEM students considering them in a STEM context. We predicted that different interpretations might lead some items to function well while others may not and that this might result in the original version of the instrument either over- or under-estimating FF in STEM contexts. Since FF has potential to influence student motivation, behavior, and even persistence in STEM, we felt that it was very important to investigate the validity of the PFAI for STEM students *in STEM contexts* and to understand how use of the original, unmodified, version of the instrument might misrepresent results.

This was a particularly important question for us to address given that our aim was to modify the PFAI to ensure accurate measurement of FF for STEM students considering STEM contexts. An accurate measure will allow future researchers to (a) determine how STEM students across contexts and from different demographic groups experience FF, (b) monitor changes in FF as a result of specific efforts or interventions designed to address it, and (c) assess what experiences lead to more or less FF in STEM. Our results support the claim that the final modified version of the PFAI is more accurate than the original unmodified version for assessing FF in STEM undergraduates.

To evaluate how measurement of FF improved as a result of iterative changes to the PFAI scale, we compared mean differences in FF across scale iterations within the same sample of students. Mean and standard error for all factors across our iterations with our Fall 2018 data including (a) Conroy et al.’s original PFAI, (b) the PFAI with STEM-specific language added (*Step 1*), and (c) our post-EFA modified version (*Step 5*) are displayed in Table 8. In addition, we present a comparison of mean differences among a subsample of our PEER data between the original PFAI model with STEM-specific language added and our fully modified model. Mean differences among these groups were tested using paired samples t-tests. In both our mixed and PEER samples, we find significant differences in the mean levels of the majority of the FF dimensions. This confirms our concern that existing assessment measures, such as the original PFAI, might misrepresent the actual level of FF experienced by undergraduates in STEM contexts. This may be a result of students not focusing their responses on STEM-specific contexts when completing the original PFAI or a result of poor fit of the factor structure of the original PFAI when considering STEM contexts. Regardless, these results support the need for the revised structure presented in this study.

**Table 8** Mean and standard error of PFAI dimensions by model type

	Complete Fall 2018 sample Mean (SE) <sup>a</sup>			PEER subsample <sup>b</sup> Mean (SE) <sup>a</sup>	
	Original Conroy model	Original model + STEM-specific language (Step 1)	Revised model; (Step 5)	Original model + STEM-specific language (Step 1)	Revised model; (Step 5)
Overall FF	2.43 (.024) <sub>A</sub>	2.47 (.023) <sub>A</sub>	2.55 (0.13) <sub>A</sub>	2.63 (0.07) <sub>A</sub>	2.69 (0.07) <sub>B</sub>
FUF	1.46 (.030) <sub>A</sub>	1.69 (.034) <sub>B</sub>	2.76 (0.09) <sub>C</sub>	2.68 (0.07) <sub>A</sub>	2.82 (0.09) <sub>B</sub>
FIOLI	2.22 (.031) <sub>A</sub>	2.12 (.030) <sub>B</sub>	2.12 (.030) <sub>B</sub>	2.08 (0.08) <sub>A</sub>	2.08 (0.08) <sub>A</sub>
FUIO	2.30 (.031) <sub>A</sub>	2.40 (.031) <sub>B</sub>	2.65 (0.12) <sub>C</sub>	2.60 (0.08) <sub>A</sub>	2.75 (0.09) <sub>B</sub>
FSE	3.30 (.030) <sub>A</sub>	3.24 (.032) <sub>A</sub>	3.18 (0.06) <sub>A</sub>	2.98 (0.09) <sub>A</sub>	3.12 (0.10) <sub>B</sub>
FDSE	2.87 (.029) <sub>A</sub>	2.88 (.032) <sub>A</sub>	N/A <sup>c</sup>	2.80 (0.08)	N/A <sup>c</sup>

Notes:<sup>a</sup> Means with differing subscripts are significantly different at  $p < .05$  based on paired samples t-tests <sup>b</sup> Approximately 60% ( $n = 166$ ) of our total PEER sample responded to the full set of the 25 original PFAI items with STEM-specific language, allowing for direct comparison of means between the original model with STEM-specific language and the revised model. <sup>c</sup>The FDSE subscale was dropped from the revised measure, so there is no descriptive information to display

The specific mean value differences within and across the FF dimensions and models yield several interesting findings. In our Fall 2018 sample, for both the original Conroy model and the version of the PFAI with STEM-specific language added, internally driven causes of failure (i.e., FSE, FDSE) were reported in the highest amounts, followed by the more immediate external causes (i.e., FUIO, FIOI), with the lowest levels of FF reported for the most distal FUF. However, when means were re-computed after the factor structure was modified via EFA, FUF became the second-highest driver of FF. Among PEER students, FUF was already a prominent worry, although significant increases were still observed with the modifications to the model. This suggests that, compared to the original PFAI, FUF seems to be especially sensitive as a motivation to avoid failure for undergraduate STEM students, particularly for PEERs. While poor exam grades are not the only challenge or failure that students face in the STEM context, they are a highly cited one. When given the opportunity to provide a recent example of a failure, approximately 50% of both samples presented in Table 8 responded with stories that involved poor scores on tests or exams. Research suggests that the perceived utility of exam results to assist with future goals is second only to perceived test difficulty in predicting student test anxiety (Bonaccio and Reeve 2010). In addition, multiple studies of premedical students found that a leading cause of STEM attrition is difficulty meeting the high demands of science courses (Lin et al. 2013). If the PFAI prompts undergraduates to specifically consider the context of their struggles and challenges within STEM, a majority may consider past failures on tests or other achievement measures. Ruminating on achievement measures that could impact their admission to graduate/medical school or other future aspirations is likely to lead to increased FUF. PEER students especially may feel like the loss of an entire career

or life goal is one failure away since they must often cope with added academic pressures from their own families and communities that result from being a minority (e.g., Robinson 2013; see also work on tokenism theory: Kanter 1977).

Another interesting change occurred among mean levels of those factors related to important others either losing interest (FIOI) or becoming upset (FUIO). While significant mean differences can be seen for both FIOI and FUIO, the direction of these differences is not consistent. Considering failures in the STEM context increased students' FUIO; however, it decreased their FIOI. Also, in both samples, students reported the lowest levels of FIOI compared to all other FF dimensions. The current generation of traditional university students express that parents are often sources of both emotional support and pressure. Specifically, many students have described the pressure they feel from parents to choose certain majors or careers and to graduate within a certain timeframe (Montag, 2012). If STEM students experience this increased parental (or other) pressure to succeed, they might expect struggles and failures to yield disappointment (higher FUIO), but not necessarily a decrease in interest (that is, interest will continue in the form of continued pressure; lower FIOI). Observing mean levels, it appears that FUIO may be reported at higher levels by PEER students, suggesting that this may be a more salient fear for minoritized students in STEM. The impact of this pressure may be particularly high for students who feel that they represent their entire family or identity group (or are made to feel this way by their instructors and peers) and thereby feel the weight of expectation and attendant disappointment in the face of any perceived struggle or failure (Kanter 1977; Robinson 2013; Robinson et al. 2013; Winkle-Wagner 2009). This is further seen in the fact that, while fears around experiencing shame and embarrassment remained the same

across models for the full Fall 2018 sample, there was a significant increase in FSE observed in the PEER subsample by using the revised model. It is also interesting to note that the FIOLI dimension was the only one which did not result in any further improvements under our EFA. It is possible that, while the modified version of the PFAI may capture unexpressed variation in the other dimensions, the FIOLI dimension already reflected a valid expression of respondents' experiences.

While the scope of this work is insufficient to draw conclusions regarding what precisely distinguishes an "effective" item from one that is less so with regard to assessing STEM-specific FF in undergraduate students, our results clearly suggest that students responding in this context do so in ways that are characteristically unique from Conroy's validation sample. Given the overall direction of these results, it appears that the original version of the PFAI is likely to significantly underestimate students' reported experiences of FF within academic STEM contexts. Future qualitative studies could explore the potential explanations proposed for variations in dimension means and seek to better understand the qualities of "good" items to assess FF in STEM contexts.

### Conclusions and future directions

Taken together, our results support a revision of Conroy's original PFAI to address STEM students' unique experiences of FF. First, students' responses appear to be STEM context-dependent. This is not surprising considering that Conroy et al. (2001) initially began classifying perceived consequences of failure with the argument that it was important to delineate such information across performance contexts. Second, STEM undergraduates show significant differences in their levels of FF based on the individual items they agree or disagree with on the PFAI. While in nearly all cases students' responses produced the same factors as the original PFAI (that is, the same types of items still clustered together during factor analysis), fewer of the items emerged for most dimensions. Clarification is needed on whether and why students feel certain items represent their fears related to failure in STEM contexts while others do not. Any basic practices or principles discovered could be extended to assessment of intrapersonal elements in STEM contexts in general.

Overall, there appears to be a pattern in our results suggesting that FSE may be reported at high levels regardless of specific context. However, FUF and FUIO are the individual PFAI dimensions reported at the highest levels specifically by STEM undergraduates while FIOLI is of overall least concern. However, this only reflects dimensions which were previously proposed and included in the PFAI. Given that our results demonstrate the

uniqueness of FF within the STEM context, research with the potential to identify other reasons STEM students may fear failure is also warranted. The purpose and scope of this work was limited to refining the PFAI for more accurate assessment within a specific context. Extensive qualitative work is recommended to elucidate whether there are additional dimensions influencing STEM students' FF.

Finally, this work underscores the point that undergraduate STEM students represent a unique population in academic achievement contexts and that accurately assessing the effects of interventions, especially those on intrapersonal elements such as FF, requires modified (or even brand new!) assessment tools. It is also especially important that we ensure our assessment tools accurately account for the experiences of nontraditional and underserved STEM students, especially PEERs, since these students are likely to leave STEM at higher rates (Asai 2020; National Science Board 2018; Steele 1997; Stinebrickner and Stinebrickner 2014). Results of our factor analyses and student interviews strongly suggest that our shorter, 15-item STEM-specific version of the PFAI provides the best fit for assessing levels of FF in STEM undergraduate students and STEM PEERs. Use of this measure will allow a more accurate assessment of FF in these populations and conveniently reduces survey burden on students.

We hope that future studies of FF will seek to bridge research with practice for STEM education improvement via the use of the modified PFAI. Research that relates levels of the various FF dimensions to STEM undergraduate academic outcomes (emotional, behavioral, or cognitive) using appropriate predictive statistical methods would be highly informative. Such work would not only help to explain and predict positive academic outcomes, but also be valuable when designing interventions to improve student success and retention in STEM majors. For interventions that target FF as a means of improving STEM student success (e.g., reducing evaluation anxiety, Hjeltnes et al. 2015; failure attribution retraining, Haynes et al. 2009), the modified PFAI could be used as a pre- and post-survey to quickly assess intervention efficacy. In particular, we would encourage researchers and practitioners using the PFAI to consider it as a tool to assess the emotional cost of engaging in research-based and active pedagogies as these contexts may *both* exacerbate FF and also help students to better cope with future failures via exposure to challenge. As instructors, we must consider not only the learning implications of incorporating active, research-based pedagogies, but also the emotional, social, and intrapersonal elements affected by our choices, especially with regard to situations in which failure is a possibility.



Future research should also investigate assessment of other intrapersonal elements, such as coping style and sense of belonging, that could be context-dependent and may require careful consideration of the population of interest. It is clear from this work and others (e.g., Knekta et al. 2019; Knekta et al. 2020; Rowland et al. 2019) that considering how to accurately assess these complex elements in STEM contexts may be imperative to gain an understanding of their roles in influencing student outcomes. Currently, there is a paucity of instruments available to measure intrapersonal elements in STEM contexts (Henry et al. 2019) and more generally a need to improve the quality of measurements for latent variables in STEM populations (Knekta et al. 2019; Limeri et al. 2020; Rowland et al. 2019). Intrapersonal elements tend to consist of latent variables that require more complex means of assessment, often construction of survey instruments with multiple items (Knekta et al. 2019). In addition, due to colloquial use of terms, intrapersonal elements are often confused with other constructs (e.g., interest and curiosity; Rowland et al. 2019) as is the case with anxiety and FF (Cacciotti et al. 2016; Lazarus 1991). Investigating, designing, and exploring validity of assessments for these elements can help to both clarify definitions while also ensuring their accurate measurement for the population in question. Beginning with existing measures valid for other populations can serve as an excellent starting point to improve upon a measure for undergraduate STEM populations (Knekta et al. 2019).

In conclusion, we recommend that researchers interested in exploring the effects of intrapersonal elements on student outcomes in STEM undergraduates make use of our modified STEM-specific PFAI assessment measure, consider revalidating other assessments of intrapersonal elements that may be context-dependent for STEM students, and continue considering best assessment practices for undergraduate STEM education research (Knekta et al. 2019). The results of this study highlight the inability of existing measures to fully capture intrapersonal elements such as FF in undergraduate STEM populations and the need for even more specific focus on assessing these elements for students from underserved groups in STEM. This underscores the continuing need to refine and develop context-specific assessment measures as we work towards a better understanding of the complex relationships between these elements and student outcomes. Only with such confidence in our assessment tools will STEM educators succeed in developing pedagogical strategies to nurture a diverse and persevering STEM workforce poised to meet and answer the complex scientific challenges of the twenty-first century.

#### Abbreviations

AIC: Akaike's Information Criterion; CFA: Confirmatory factor analysis; CFI: Comparative Fit Index; EFA: Exploratory factor analysis; FF: Fear of failure;

FDSE: Fear of devaluing one's self-estimate; FIOLI: Fear of important others losing interest; FSE: Fear of shame or embarrassment; FUF: Fear of having an uncertain future; FUIO: Fear of upsetting important others; MCAR: Missing completely at random; MLR: Maximum likelihood ratio; N: Sample size; NSF: National Science Foundation; PEERS: Persons excluded because of their ethnicity or race; PFAI: Performance Failure Appraisal Inventory; RMSEA: Root Mean Square Error Approximation; SES: Socioeconomic status; SRMR: Standardized mean square residual; STEM: Science, Technology, Engineering, and Mathematics

## Supplementary Informations

The online version contains supplementary material available at <https://doi.org/10.1186/s40594-021-00300-4>.

### Additional file 1 .

#### Acknowledgements

We acknowledge members of *Factors affecting Learning, Attitudes, and Mindsets in Education network* (FLAMENet), the Heemstra Lab, and the RE<sup>3</sup>ACH Lab for their support in reading and commenting on early drafts of this work. We also sincerely thank the many STEM undergraduate students and instructors who participated in this research. We acknowledge support from the National Science Foundation (RCN UBE Incubator 1827160; RCN UBE 1919953).

#### Authors' contributions

MAH led the study. MAH worked closely with other partners to conceptualize the project and then led participant recruitment; data collection, cleaning, and curation; data analysis and interpretation; manuscript drafting; and manuscript revisions and editing. SS contributed to conceptualization of the project and assisted with participant recruitment and data collection. SS also assisted with background research and literature review. LKC and JMH contributed to conceptualization of the project, provided guidance on project design and next steps throughout the project, and contributed to subject recruitment. LKC and JMH contributed to manuscript drafting with extensive revisions and comments for improvement and provided critical perspectives as STEM educators, practitioners, and administrators. BJL provided advice and guidance on experimental design and specific advice on statistical approaches and analyses. BJL provided extensive consultation on quantitative methodologies that were most appropriate especially when troubleshooting was necessary. LAC led conceptualization of the work with MAH and worked closely with MAH and BLC to design the study methodology and determine statistical approaches for the analysis. LAC provided assistance and consultation in all areas of the study including recruitment, data collection, analysis, and data interpretation. LAC worked with MAH to draft the manuscript and collaborated in revising and editing the final manuscript. The authors read and approved the final manuscript.

#### Funding

This work was supported by two National Science Foundation, Improving Undergraduate STEM Education (IUSE) Research Coordination Network (RCN) for Undergraduate Biology Education (UBE) grants (RCN UBE Incubator 1827160; RCN UBE 1919953)

#### Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to the potential to indirectly identify individual study participants based on a combination of demographic characteristics and location data. However, select subsets of data are available from the corresponding author on reasonable request.

#### Declarations

#### Competing interests

The authors declare that they have no competing interests

**Author details**<sup>1</sup>Department of Chemistry, Emory University, Atlanta, GA 30322, USA.<sup>2</sup>Department of Chemistry, Haverford College, Haverford, PA 19041, USA.<sup>3</sup>Department of Psychology, Haverford College, Haverford, PA 19041, USA.<sup>4</sup>Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, 1900 Pleasant St., 334 UCB, Boulder, CO 80309, USA.

Received: 22 December 2020 Accepted: 7 June 2021

Published online: 06 July 2021

**References**

- Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of Hirotugu Akaike* (pp. 199–213). New York, NY: Springer.
- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, *38*(2), 113–125. <https://doi.org/10.1006/jesp.2001.1491>.
- Asai, D. J. (2020). Race matters. *Cell*, *181*(4), 754–757. <https://doi.org/10.1016/j.cell.2020.03.044>.
- Asparouhov, T., & Muthén, B. (2018). SRMR in Mplus. Retrieved from Mplus Web Notes website: <http://www.statmodel.com/download/SRMR2.pdf>.
- Auchincloss, L., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I., ... Dolan, E. L. (2014). Assessment of course-based undergraduate research experiences: A meeting report. *CBE—Life Sciences Education*, *13*(1), 29–40. <https://doi.org/10.1187/cbe.14-01-0004>.
- Bandura, A., Freeman, W. H., & Lightsey, R. (1999). *Self-efficacy: The exercise of control*.
- Barbosa, P. (1975). Commentary: underrepresentation of minorities in the biological sciences. *Bioscience*, *25*(5), 319–320. <https://doi.org/10.2307/1297130>.
- Bartels, J. M., & Herman, W. E. (2011, May). *Fear of failure, self-handicapping, and negative emotions in response to fear of failure*. Poster presented at: 23rd Annual Convention for the Association for Psychological Science (Washington, DC).
- Berglas, S., & Jones, E. E. (1978). Drug choice as a self-handicapping strategy in response to noncontingent success. *Journal of Personality and Social Psychology*, *36*(4), 405–417. <https://doi.org/10.1037/0022-3514.36.4.405>.
- Bledsoe, T., & Baskin, J. (2014). Recognizing student fear: The elephant in the classroom. *College Teaching*, *62*(1), 32–41. <https://doi.org/10.1080/87567555.2013.831022>.
- Bonaccio, S., & Reeve, C. L. (2010). The nature and relative importance of students' perceptions of the sources of test anxiety. *Learning and Individual Differences*, *20*(6), 617–625. <https://doi.org/10.1016/j.lindif.2010.09.007>.
- Boomsma, A., & Hoogland, J. J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural Equation Models: Present and Future. A Festschrift in Honor of Karl Jöreskog*, (pp. 139–168). Scientific Software International: Lincolnwood, IL.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*, (2nd ed., ). New York, NY: Guilford Press.
- Cacciotti, G. (2015). *Fear of failure in entrepreneurship: a review, reconceptualization, and operationalization* (doctoral dissertation). [http://wrap.warwick.ac.uk/73258/1/WRAP\\_THESIS\\_Cacciotti\\_2015.pdf](http://wrap.warwick.ac.uk/73258/1/WRAP_THESIS_Cacciotti_2015.pdf). Accessed 18 Dec 2020.
- Cacciotti, G., Hayton, J. C., Mitchell, J. R., & Giazitoglu, A. (2016). A reconceptualization of fear of failure in entrepreneurship. *Journal of Business Venturing*, *31*(3), 302–325. <https://doi.org/10.1016/j.jbusvent.2016.02.002>.
- Cape, P., & Phillips, K. (2015). Questionnaire length and fatigue effects: the latest thinking and practical solutions. White paper. Available online at: [www.surveysampling.com/site/assets/files/1586/questionnaire-length-and-fatigue-effects-the-latest-thinking-and-practical-solutions.pdf](http://www.surveysampling.com/site/assets/files/1586/questionnaire-length-and-fatigue-effects-the-latest-thinking-and-practical-solutions.pdf). Accessed 31 July 2017.
- Caraway, K., Tucker, C. M., Reinke, W. M., & Hall, C. (2003). Self-efficacy, goal orientation, and fear of failure as predictors of school engagement in high school students. *Psychology in the Schools*, *40*(4), 417–427. <https://doi.org/10.1002/pits.10092>.
- Carver, C. S. (1997). You want to measure coping but your protocol's too long: consider the briefCOPE. *International Journal of Behavioral Medicine*, *4*(1), 92.
- Cattell, R. B. (1966). The Scree Test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276. [https://doi.org/10.1207/s15327906mbr0102\\_10](https://doi.org/10.1207/s15327906mbr0102_10).
- Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York: Plenum Press. <https://doi.org/10.1007/978-1-4684-2262-7>.
- Ceyhan, G. D., & Tillotson, J. W. (2020). Early year undergraduate researchers' reflections on the values and perceived costs of their research experience. *International Journal of STEM Education*, *7*(1), 1–19.
- Chang, M. J., Sharkness, J., Hurtado, S., & Newman, C. B. (2014). What matters in college for retaining aspiring scientists and engineers from underrepresented racial groups. *Journal of Research in Science Teaching*, *51*(5), 555–580. <https://doi.org/10.1002/tea.21146>.
- Chen, L. H., Wu, C. H., Kee, Y. H., Lin, M. S., & Shui, S. H. (2009). Fear of failure, 2 X 2 achievement goal and self-handicapping: an examination of the hierarchical model of achievement motivation in physical education. *Contemporary Educational Psychology*, *34*(4), 298–305. <https://doi.org/10.1016/j.cedpsych.2009.06.006>.
- Conroy, D. E. (2001). Progress in the development of a multidimensional measure of fear of failure: The performance failure appraisal inventory (PFAI). *Anxiety, Stress, and Coping*, *14*(4), 431–452. <https://doi.org/10.1080/10615800108248365>.
- Conroy, D. E., Poczwadowski, A., & Henschen, K. P. (2001). Evaluative criteria and consequences associated with failure and success for elite athletes and performing artists. *Journal of Applied Sports Psychology*, *13*(3), 300–322. <https://doi.org/10.1080/104132001753144428>.
- Conroy, D. E., Willow, J. P., & Metzler, J. N. (2002). Multidimensional fear of failure measurement: the Performance Failure Appraisal Inventory. *Journal of Applied Sport Psychology*, *14*(2), 76–90. <https://doi.org/10.1080/10413200252907752>.
- Cooper, K. M., Downing, V. R., & Brownell, S. E. (2018). The influence of active learning practices on student anxiety in large-enrollment college science classrooms. *International Journal of STEM Education*, *5*(1), 1–18.
- Corwin, L. A., Graham, M. J., & Dolan, E. L. (2015). Modeling course-based undergraduate research experiences: An agenda for future research and evaluation. *CBE—Life Sciences Education*, *14*(1), e51.
- Corwin, L. A., Morton, T. R., Demetriou, C., & Panter, A. T. (2020). A qualitative investigation of STEM students' switch to non-STEM majors post-transfer. *Journal of Women and Minorities in Science and Engineering*, *26*(3), 263–301. <https://doi.org/10.1615/JWomenMinorScienEng.2020027736>.
- Cox, R. D. (2009). "It's just that I was afraid": Promoting success by addressing students' fear of failure. *Community College Review*, *37*(1), 52–81. <https://doi.org/10.1177/0091552109338390>.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. <https://doi.org/10.1007/BF02310555>.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. <https://doi.org/10.1037/h0040957>.
- Cullinane, J. (2009). Diversifying the STEM Pipeline: the model replication institutions program. Institute for Higher Education Policy. <https://files.eric.ed.gov/fulltext/ED508104.pdf>. Accessed 18 Dec 2020.
- De Castella, K., Byrne, D., & Covington, M. (2013). Unmotivated or motivated to fail? A cross-cultural study of achievement motivation, fear of failure, and student disengagement. *Journal of Educational Psychology*, *105*(3), 861–880. <https://doi.org/10.1037/a0032464>.
- Downing, V. R., Cooper, K. M., Cala, J. M., Gin, L. E., & Brownell, S. E. (2020). Fear of negative evaluation and student anxiety in community college active-learning science courses. *CBE—Life Sciences Education*, *19*(2), ar20.
- Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Ballantine.
- Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, *72*(1), 218–232. <https://doi.org/10.1037/0022-3514.72.1.218>.
- Elliot, A. J., & Church, M. A. (2003). A motivational analysis of defensive pessimism and self-handicapping. *Journal of Personality*, *71*(3), 369–396. <https://doi.org/10.1111/1467-6494.7103005>.
- Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, *80*(3), 501–519. <https://doi.org/10.1037/0022-3514.80.3.501>.
- Elliot, A. J., & Thrash, T. M. (2004). The intergenerational transmission of fear of failure. *Personality and Social Psychology Bulletin*, *30*(8), 957–971. <https://doi.org/10.1177/0146167203262024>.
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 1–11. <https://doi.org/10.3758/BF03203630>.
- Estrada, M., Burnett, M., Campbell, A. G., Campbell, P. B., Denetclaw, W. F., Gutiérrez, C. G., ... Zavala, M. E. (2016). Improving underrepresented minority student persistence in STEM. *CBE—Life Sciences Education*, *15*(3), e5.
- Estrada, M., Hernandez, P. R., & Schultz, P. W. (2018). A longitudinal study of how quality mentorship and research experience integrate underrepresented minorities into STEM careers. *CBE—Life Sciences Education*, *17*(1), ar9.

- Estrada, M., & Matsui, J. (2019). A longitudinal study of the biology scholars' program: Maintaining student integration and intention to persist in science career pathways. *Understanding Interventions*, 10(1), 9884.
- Estrada, M., Woodcock, A., Hernandez, P. R., & Schultz, P. W. (2011). Toward a model of social influence that explains minority student integration into the scientific community. *Journal of Educational Psychology*, 103(1), 206–222. <https://doi.org/10.1037/a0020743>.
- Farrington, C.A. (2019). *Noncognitive outcomes of liberal arts education*. The Andrew W. Mellon Foundation. <https://mellon.org/news-blog/articles/noncognitive-factors-college-experience/> Accessed 18 Dec 2020.
- Fink, A., Cahill, M. J., McDaniel, M. A., Hoffman, A., & Frey, R. F. (2018). Improving general chemistry performance through a growth mindset intervention: Selective effects on underrepresented minorities. *Chemistry Education Research and Practice*, 19(3), 783–806. <https://doi.org/10.1039/c7rp00244k>.
- Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, 20(2), 303–315. <https://doi.org/10.1086/209351>.
- Freeman, R. B., & Huang, W. (2014). Collaboration: Strength in diversity. *Nature News*, 513(7518), 305. <https://doi.org/10.1038/513305a>.
- Gasiewski, J. A., Eagan, M. K., Garcia, G. A., Hurtado, S., & Chang, M. J. (2012). From gatekeeping to engagement: A multicontextual, mixed method study of student academic engagement in introductory STEM courses. *Research in Higher Education*, 53(2), 229–261. <https://doi.org/10.1007/s11162-011-9247-y>.
- George, D., & Mallery, P. (14th Ed.). (2016). *IBM SPSS Statistics 23 Step by Step: A Simple Guide and Reference*. USA: Routledge, DOI: <https://doi.org/10.4324/9781315545899>.
- Gin, L. E., Rowland, A. A., Steinwand, B., Bruno, J., & Corwin, L. A. (2018). Students who fail to achieve predefined research goals may still experience many positive outcomes as a result of CURE participation. *CBE—Life Sciences Education*, 17(4), ar57.
- Gnamb, T., & Kaspar, K. (2014). Disclosure of sensitive behaviors across self-administered survey modes: A meta-analysis. *Behavior Research Methods*, 47(4), 1237–1259.
- Hacker, K. (2013). *Community-based participatory research*. New York, NY: Sage publications. <https://doi.org/10.4135/9781452244181>.
- Harsh, J. A., Maltese, A. V., & Tai, R. H. (2011). Undergraduate research experiences from a longitudinal perspective. *Journal of College Science Teaching*, 41(1), 84–91.
- Haynes, T. L., Perry, R. P., Stupnisky, R. H., & Daniels, L. M. (2009). *A review of attributional retraining treatments: Fostering engagement and persistence in vulnerable college students*. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research*, (pp. 227–272). Dordrecht: Springer. [https://doi.org/10.1007/978-1-4020-9628-0\\_6](https://doi.org/10.1007/978-1-4020-9628-0_6).
- Henry, M. A., Shorter, S., Charkoudian, L., Heemstra, J. M., & Corwin, L. A. (2019). FAIL is not a four-letter word: a theoretical framework for exploring undergraduate students' approaches to academic challenges and responses to failure in STEM learning environments. *CBE—Life Sciences Education*, 18(ar11), 1–17.
- Hjeltnes, A., Binder, P. E., Moltu, C., & Dundas, I. (2015). Facing the fear of failure: an explorative qualitative study of client experiences in a mindfulness-based stress reduction program for university students with academic evaluation anxiety. *International Journal of Qualitative Studies on Health and Well-Being*, 10(1), 27990. <https://doi.org/10.3402/qhw.v10.27990>.
- Hoaglin, D. C., & Iglewicz, B. (1987). Fine tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 82(400), 1147–1149. <https://doi.org/10.1080/01621459.1987.10478551>.
- Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81(396), 991–999. <https://doi.org/10.1080/01621459.1986.10478363>.
- Hurtado, S., Newman, C. B., Tran, M. C., & Chang, M. J. (2010). Improving the rate of success for underrepresented racial minorities in STEM fields: Insights from a national project. *New Directions for Institutional Research*, 2010(148), 5–15. <https://doi.org/10.1002/ir.357>.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141–151. <https://doi.org/10.1177/001316446002000116>.
- Kanter, R. M. (1977). Some effects of proportions on group life: skewed sex ratios and responses to token women. *The American Journal of Sociology*, 82(5), 965–990. <https://doi.org/10.1086/226425>.
- Kenny, D. A. (2020, June 5). *Measuring model fit*. Retrieved December 18, 2020 from <https://davidakenny.net/cm/fit.htm>.
- Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44(3), 486–507. <https://doi.org/10.1177/0049124114543236>.
- Kline, R. B. (2010). *Principles and Practice of Structural Equation Modeling*, (3rd ed., ). New York, NY: Guilford Press.
- Knekta, E., Rowland, A. A., Corwin, L. A., & Eddy, S. (2020). Measuring university students' interest in biology: Evaluation of an instrument targeting Hidi and Renninger's individual interest. *International Journal of STEM Education*, 7, 1–16.
- Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: using factor analysis to gather validity evidence when using surveys in your research. *CBE—Life Sciences Education*, 18(rm1), 1–17.
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>.
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, 47(4), 2025–2047. <https://doi.org/10.1007/s11135-011-9640-9>.
- Kuh, G. D., Kinzie, J. L., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2006). *What matters to student success: a review of the literature* (Vol. 8). Washington, DC: National Postsecondary Education Cooperative.
- Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, 9(8), 2207–2230. <https://doi.org/10.4236/psych.2018.98126>.
- Laursen, S., Hunter, A.-B., Seymour, E., Thiry, H., & Melton, G. (2010). *Undergraduate research in the sciences: engaging students in real science*. Hoboken, NJ: John Wiley & Sons Inc.
- Lazarus, R. S. (1991). *Emotion and Adaptation*. New York, NY: Oxford University Press.
- Lazarus, R. S. (1993). Coping theory and research: Past, present, and future. In R. S. Lazarus (Ed.), *Fifty years of the research and theory of RS Lazarus: an analysis of historical and perennial issues*, (pp. 366–388). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, 347(6219), 262–265. <https://doi.org/10.1126/science.1261375>.
- Limeri, L. B., Choe, J., Harper, H. G., Martin, H. R., Benton, A., & Dolan, E. L. (2020). Knowledge or abilities? How undergraduates define intelligence. *CBE—Life Sciences Education*, 19(1), ar5.
- Lin, K. Y., Parnami, S., Fuhrel-Forbis, A., Anspach, R. R., Crawford, B., & De Vries, R. G. (2013). The undergraduate premedical experience in the United States: a critical review. *International Journal of Medical Education*, 4, 26–37. <https://doi.org/10.5116/ijme.5103.a8d3>.
- Little, T. D. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404), 1198–1202. <https://doi.org/10.1080/01621459.1988.10478722>.
- Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Publications, Inc.
- Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives*, 7(4), 199–204. <https://doi.org/10.1111/cdep.12043>.
- Lopatko, D., Alvarez, C., Bernard, D., & Chandrasekaran, C. (2008). Undergraduate research: Genomics education partnership, *Science*, 322(5902), 684–685, doi: <http://doi.org/https://doi.org/10.1126/science.1165351>.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130–149. <https://doi.org/10.1037/1082-989X.1.2.130>.
- Marsh, H. W., & Hau, K. T. (1999). Confirmatory factor analysis: Strategies for small sample sizes. *Statistical Strategies for Small Sample Research*, 1, 251–284.
- Martin, A. J., & Marsh, H. W. (2003). Fear of failure: Friend or foe? *Australian Psychologist*, 38(1), 31–38. <https://doi.org/10.1080/00050060310001706997>.
- Masaki, M. (2010). How to factor-analyze your data right: Do's, don't's, and how to's. *International Journal of Psychological Research*, 3(1), 97–110.
- Maton, K. I., Domingo, M. R. S., Stolle-McAllister, K. E., Zimmerman, J. L., & Hrabowski III, F. A. (2009). Enhancing the number of African Americans who pursue STEM PhDs: Meyerhoff Scholarship Program outcomes, processes, and individual predictors, *Journal of Women and Minorities in Science and Engineering*, 15(1), 15.
- Matthews, C. M. (1990). Underrepresented Minorities and Women in Science, Mathematics, and Engineering: Problems and Issues for the 1990s. CRS Report for Congress. <https://files.eric.ed.gov/fulltext/ED337525.pdf> Accessed 18 Dec 2020.
- Muthén, L. K., & Muthén, B. O. (1998-2018). *Mplus User's Guide*, (Eighth ed., ). Los Angeles, CA: Muthén & Muthén.

- National Research Council (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, D.C.: National Academies Press.
- National Science Board, *Science and Engineering Indicators 2018* (National Science Foundation, 2018).
- National Science Foundation. (2020). Science and engineering degrees, by race and ethnicity of recipients: 2008-2018. [Data file]. <https://www.nsf.gov/statistics/degreerecipients/>. Accessed 18 Dec 2020.
- Nelson, K. L., Nelson, K. K., McDaniel, J. R., & Tackett, S. (2019). Majoring in STEM: How the factors of fear of failure, impostor phenomenon, and self-efficacy impact decision-making. *National Social Science Journal*, 52(1), 79–82.
- Noguera, M., Alvarez, C., & Urbano, D. (2013). Socio-cultural factors and female entrepreneurship. *International Entrepreneurship and Management Journal*, 9(2), 183–197. <https://doi.org/10.1007/s11365-013-0251-x>.
- Onwuegbuzie, A. J. (2004). Academic procrastination and statistics anxiety. *Assessment & Evaluation in Higher Education*, 29(1), 3–19. <https://doi.org/10.1080/0260293042000160384>.
- Pajares, F. (2005). *Gender differences in mathematics self-efficacy beliefs*. Cambridge, UK: Cambridge University Press.
- Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco, CA: Jossey-Bass Inc.
- Pelin, M. E. T. E., & Subasi, M. (2020). The relationship between academic coping, approach achievement goals and the fear of shame and embarrassment in science class. *Journal of Education in Science Environment and Health*, 7(1), 15–25.
- Peytchev, A., & Peytcheva, E. (2017). Reduction of measurement error due to survey length: Evaluation of the split questionnaire design approach. *Survey Research Methods*, 11(4), 361–368.
- Revilla, M., & Ochoa, C. (2017). Ideal and maximum length for a web survey. *International Journal of Market Research*, 59(5), 557–565.
- Rhemtulla, M., & Little, T. (2012). Tools of the trade: planned missing data designs for research in cognitive development. *Journal of Cognitive Development*, 13(4), 1–12.
- Robinson, S. J. (2013). Spoke tokenism: Black women talking back about graduate school experiences. *Race Ethnicity and Education*, 16(2), 155–181. <https://doi.org/10.1080/13613324.2011.645567>.
- Robinson, S. J., Esquibel, E., & Rich, M. D. (2013). "I'm Still Here:" Black female undergraduates' self-definition narratives. *World Journal of Education*, 3(5), 57–71.
- Rowland, A. A., Knekta, E., Eddy, S., & Corwin, L. A. (2019). Defining and measuring students' interest in biology: An analysis of the biology education literature. *CBE—Life Sciences Education*, 18(3), ar34.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York, NY: Wiley. <https://doi.org/10.1002/9780470316696>.
- Sagar, S., Busch, B. K., & Jowet, S. (2010). Success and failure, fear of failure, and coping response of adolescent academy football players. *Journal of Applied Sport Psychology*, 2(2), 213–220.
- Schäfer, J., Reid, N., Cox, D., Keiding, N., Louis, T., Tong, H., & Isham, V. (1997). *Analysis of incomplete multivariate data*. New York, NY: Chapman and Hall. <https://doi.org/10.1201/9781439821862>.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54(2), 93–105. <https://doi.org/10.1037/0003-066X.54.2.93>.
- Seymour, E., & Hunter, A. B. (2019). *Talking About Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education*. Cham, Switzerland: Springer. <https://doi.org/10.1007/978-3-030-25304-2>.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>.
- Simpson, A., & Maltese, A. (2017). "Failure is a major component of learning anything": the role of failure in the development of STEM professionals. *Journal of Science Education and Technology*, 26(2), 223–237. <https://doi.org/10.1007/s10956-016-9674-9>.
- Sisk, V. F., Burgoyne, A. P., Sun, J., Butler, J. L., & Macnamara, B. N. (2018). To what extent and under which circumstances are growth mind-sets important to academic achievement? Two meta-analyses. *Psychological Science*, 29(4), 549–571. <https://doi.org/10.1177/0956797617739704>.
- Skinner, E. A., Edge, K., Altman, J., & Sherwood, H. (2003). Searching for the structure of coping: a review and critique of category systems for classifying ways of coping. *Psychological Bulletin*, 129(2), 216–269. <https://doi.org/10.1037/0033-2909.129.2.216>.
- Snibbe, A. C., & Markus, H. R. (2005). You can't always get what you want: Educational attainment, agency, and choice. *Journal of Personality and Social Psychology*, 88(4), 703–720. <https://doi.org/10.1037/0022-3514.88.4.703>.
- Steele, C. M. (1997). A threat in the air. How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), 613–629. <https://doi.org/10.1037/0003-066X.52.6.613>.
- Stinebrickner, R., & Stinebrickner, T. (2014). Academic performance and college dropout: using longitudinal expectations data to estimate a learning model. *Journal of Labor Economics*, 32(3), 601–644. <https://doi.org/10.1086/675308>.
- Storage, D., Horne, Z., Cimpian, A., & Leslie, S. J. (2016). The frequency of "brilliant" and "genius" in teaching evaluations predicts the representation of women and African Americans across fields. *PLoS one*, 11(3), e0150194. <https://doi.org/10.1371/journal.pone.0150194>.
- Taasoobshirazi, G., & Wang, S. (2016). The performance of the SRMR, RMSEA, CFI, and TL: An examination of sample size, path size, and degrees of freedom. *Journal of Applied Quantitative Methods*, 13(1), 31–39.
- Thiry, H., Weston, T. J., Laursen, S. L., & Hunter, A. B. (2012). The benefits of multi-year research experiences: differences in novice and experienced students' reported gains from undergraduate research. *CBE—Life Sciences Education*, 11(3), 260–272. doi: <http://doi.org/https://doi.org/10.1187/cbe.11-11-0098>, 11, 3, 260, 272
- Traphagen, S. (2015, May 13). Teacher: The important conversations we are too 'scared' to have. *The Washington Post*. Retrieved December 18, 2020 from [https://www.washingtonpost.com/news/answer-sheet/wp/2015/05/13/teacher-the-helpful-conversations-we-are-too-scared-to-have/?utm\\_term=.6dc22b2d070c](https://www.washingtonpost.com/news/answer-sheet/wp/2015/05/13/teacher-the-helpful-conversations-we-are-too-scared-to-have/?utm_term=.6dc22b2d070c).
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- U.S. Department of Education, Institute of Education Sciences (2012). STEM in postsecondary education: Entrance, attrition, and coursetaking among 2003–04 beginning postsecondary students. <https://files.eric.ed.gov/fulltext/ED566425.pdf>. Accessed 18 Dec 2020.
- Watkins, M. W. (2018). Exploratory factor analysis: a guide to best practices. *Journal of Black Psychology*, 44(3), 219–246. <https://doi.org/10.1177/0095798418771807>.
- Willis, G. B. (2015). *Analysis of the cognitive interview in questionnaire design*. Oxford, UK: Oxford University Press.
- Winkle-Wagner, R. (2009). *The unchosen me: Race, gender, and identity among Black women in college*. Baltimore, MD: JHU Press.
- Yeager, D. S., Walton, G. M., Brady, S. T., Akcinar, E. N., Paunesku, D., Keene, L., ... Dweck, C. S. (2016). Teaching a lay theory before college narrows achievement gaps at scale. *Proceedings of the National Academy of Sciences*, 113(24), E3341–E3348. <https://doi.org/10.1073/pnas.1524360113>.
- Zhang, Y., Dong, S., Fang, W., Chai, X., Mei, J., & Fan, X. (2018). Self-efficacy for self-regulation and fear of failure as mediators between self-esteem and academic procrastination among undergraduates in health professions. *Adv Health Sci Educ Theory Pract*, 23(4), 817–830.
- Zuckerman, M., & Tsai, F. F. (2005). Costs of self-handicapping. *Journal of Personality*, 73(2), 411–442. <https://doi.org/10.1111/j.1467-6494.2005.00314.x>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)