

NAEP Technical Advisory Panel Proceedings of the Symposium on Academic Preparedness Research

Prepared for: National Assessment Governing Board
800 North Capitol Street, N.W., Suite 825
Washington, DC 20002 42330

Authors: Rodney A. McCloy
Tiffany C. Day

Prepared under: National Assessment of Educational Progress (NAEP)
Reading and Mathematics Frameworks: 12th Grade
Preparedness
ED NAG 13 C 0001

Date: January 15, 2015

NAEP Technical Advisory Panel Symposium Proceedings

Table of Contents

Meeting Summary.....	1
Attendees.....	1
Meeting Materials.....	2
Meeting Overview	2
Overview of NAEP Activities	2
Summaries of Paper Discussions.....	8
Using 8 th and 12 th Grade NAEP to Measure Student Readiness for Careers Barbara Plake, University of Nebraska – Lincoln.....	8
Grit: A Useful Concept in College and Career Preparedness? Ann Marie Ryan, Michigan State University.....	10
Relating NAEP to Commercial Off-the-Shelf Measures Nancy Tippins, Corporate Executive Board – Valtera Corporation	13
Group Discussion.....	15
Next Steps	16
References	18
Appendix A Meeting Agenda.....	A-1
Appendix B Update on NAEP and Academic Preparedness for College and Job Training (Presentation)	B-1
Appendix C Using 8 th and 12 th Grade NAEP to Measure Student Readiness for Careers (Paper).....	C-1
Appendix D Using 8 th and 12 th Grade NAEP to Measure Student Readiness for Careers (Presentation)	D-1
Appendix E Grit: A Useful Concept in College and Career Preparedness? (Paper).....	E-1
Appendix F Grit: A Useful Concept in College and Career Preparedness? (Presentation)	F-1
Appendix G Relating NAEP to Commercial Off-the-shelf Measures (Paper)	G-1
Appendix H Relating NAEP to Commercial Off-the-Shelf Measures (Presentation).....	H-1

**National Assessment Governing Board
NAEP Framework Evaluation Project
Technical Advisory Panel (TAP) Symposium
August 20, 2014**

MEETING SUMMARY

The Human Resources Research Organization (HumRRO) convened a meeting of the National Assessment of Educational Progress (NAEP) Framework Evaluation Technical Advisory Panel (TAP) on August 20, 2014 in Washington, DC. Prior to this meeting, Governing Board staff reviewed proposals submitted by TAP panelists and commissioned papers to be completed by the panelists. In this meeting, authors presented final papers and the panelists discussed them.

ATTENDEES

Technical Advisory Panel (TAP) Members

Michael A. Campion, Purdue University
Gregory J. Cizek, University of North Carolina, Chapel Hill
Brian Gong, Center for Assessment, National Center for the Improvement of Educational Assessment
Ronald K. Hambleton, University of Massachusetts, Amherst
Barbara S. Plake, University of Nebraska – Lincoln, retired, private consultant
Ann Marie Ryan, Michigan State University
Nancy T. Tippins, Corporate Executive Board – Valtera Corporation

National Assessment Governing Board (NAGB)

Mary Crovo, Deputy Executive Director
Sharyn Rosenberg, Assistant Director for Psychometrics
Michelle Blair, Senior Research Associate

National Center for Education Statistics (NCES)

Samantha Burg, Assessments Division, Reporting and Dissemination Branch

HumRRO

Rod McCloy, Technical Advisory Panel Task Leader
Arthur Thacker, Project Director
Andrea Sinclair, Deputy Project Director
Sunny Becker, Project Manager
Lauress (“Laurie”) Wise, Key Advisor

CRP, Inc.

Monica Duda

Tiffany Day (HumRRO) and Jennifer Ragins (CRP) prepared symposium notes.

MEETING MATERIALS

Several appendices are attached to these notes:

- Appendix A. Meeting Agenda
- Appendix B. POWERPOINT: Update on NAEP and Academic Preparedness for College and Job Training, Sharyn Rosenberg – NAGB
- Appendix C. PAPER: *Using 8th and 12th Grade NAEP to Measure Student Readiness for Careers*, Barbara Plake, University of Nebraska – Lincoln
- Appendix D. POWERPOINT: Using 8th and 12th Grade NAEP to Measure Student Readiness for Careers
- Appendix E. PAPER: *Grit: A Useful Concept in College and Career Preparedness?* Ann Marie Ryan, Michigan State University
- Appendix F. POWERPOINT: Grit: A Useful Concept in College and Career Preparedness?
- Appendix G. PAPER: *Relating NAEP to Commercial Off-the-Shelf Measures*, Nancy Tippins, Corporate Executive Board – Valtera Corporation
- Appendix H. POWERPOINT: Relating NAEP to Commercial Off-the-Shelf Measures

MEETING OVERVIEW

Dr. McCloy welcomed the panel members, Governing Board staff, and HumRRO and described the primary purpose of the meeting – to have authors present their research ideas on topics that might be brought into the NAEP preparedness research and to engage the panelists in discussion to determine the utility of NAEP as an indicator of academic preparedness for entry into college and job training. This meeting is one task in a larger project, *Evaluating Reading and Mathematics Frameworks and Item Pools as Measures of Academic Preparedness for College and Job Training*.

Specific goals for the meeting included:

1. Discussing updates on current 12th grade NAEP preparedness research.
2. Presenting and discussing innovative research ideas related to the area of preparedness, with a focus on job-training preparedness.
3. Encouraging TAP discussion on the paper presentations and other technical issues related to NAEP preparedness research.

OVERVIEW OF NAEP ACTIVITIES

Update on NAEP and Academic Preparedness for College and Job Training

Dr. Rosenberg provided an updated version of her October 10, 2013 presentation on the Governing Board's work on 12th grade preparedness (**see Appendix B**). The bulk of the presentation focused on academic preparedness for college, as this is where the majority of the research has been focused to date.

NAEP has been conducting work on preparedness for more than a decade. In the past, questions were raised as to whether 12th grade results were valid and useful, particularly given the sharp declines in school and student participation from 1988 to 2002. The Governing Board convened a 12th Grade Commission in 2003–2004 to review the current purpose, strengths, and weaknesses of 12th grade assessment and reporting. One of the Commission’s key recommendations was to turn 12th grade NAEP into a measure of academic preparedness, focusing on reading and mathematics skills and not implying other skills and traits that might be essential for full “readiness.”

In an overview, Dr. Rosenberg discussed how a significant percentage of high school graduates are not prepared for entry-level college courses. Every year, high school graduates discover that, despite being eligible to attend college, they are not well prepared for postsecondary studies. Graduates are entering college assuming they have met prerequisites. However, research suggests that a high percentage of college freshmen are required to take remedial courses before registering for credit-bearing courses. This raises the concern that high school completion does not necessarily translate into college readiness.

Since 2010, the Governing Board has been conducting research studies in five areas (content comparison; statistical relationships; judgmental standard setting; survey of postsecondary institutions; benchmarking) to determine if the nation’s 12th graders are well prepared academically for college and job training. Dr. Rosenberg provided background information and provided a high-level summary of the research conducted so far, as well as projected plans for further 12th grade preparedness research.

Despite the Governing Board’s efforts, this research has faced many difficulties. The U.S. does not have a common measure or definition of 12th grade preparedness. Postsecondary education and training in the U.S. is a multifaceted enterprise of public, private and proprietary institutions and organizations. College entrance requirements vary across institutions (and even within institutions, depending on the course and major). There are different admissions requirements and policies for remediation and placement. In addition, postsecondary job training is very complex, and there is no typical process for preparing high school graduates for entrance into the world of work. Due to that lack of uniformity, there is no single definition for preparedness for college or job training. Consequently, a goal of the Governing Board’s 12th Grade Preparedness Research Program was to craft a definition of preparedness to design and conduct the NAEP research to answer the question: Are the nation’s 12th graders prepared academically for college and job training?

The Governing Board did not assume that skills needed for college and job training were the same but did assume that students should leave high school with the ability to pursue any goal they choose, whether college or job training. (Preparedness for military entrance and job training should be included; however, the Governing Board has been unable to secure military entrance test and job-training data.)

College Preparedness

The Governing Board’s early work on this topic involved crafting a working definition of academic preparedness for college: “The knowledge and skills to qualify for placement into entry-level, credit-bearing college courses without a need for remediation in math or reading.”

Dr. Rosenberg noted that NAEP measures academic preparedness.¹ NAEP's content is similar to the college admissions examinations, according to Dr. Rosenberg. While the ACT, SAT, and ACCUPLACER content is generally covered by NAEP, NAEP's content is broader, including more (a) types of questions, (b) cognitive rigor of test questions, (c) types of reading texts, and (d) weighting and coverage of specific content in mathematics.

Key Takeaways: Academic Preparedness for College

Key takeaways from the academic preparedness research for college included the following:

- Content comparisons confirm the appropriateness of the recently revised NAEP reading and mathematics frameworks.
- Students scoring at or above proficient are likely to be academically prepared.
- However, results are different for reading and math in relation to the NAEP proficient achievement level and interpretations and implications for reporting need to be considered.

The Governing Board voted at its August 2013 meeting on a motion to approve the following statements for reporting the 2013 NAEP 12th grade reading and mathematics assessments:

“Given the design, content, and characteristics of the NAEP 12th grade reading assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness:

- The percentage of students scoring at or above a score of 302 on grade 12 NAEP in reading is a plausible estimate of the percentage of students who possess the knowledge, skills, and abilities in reading, making them academically prepared for college.
- A score of 302 corresponds to the cut score for the proficient achievement level in 12th grade Reading.”

In 2013, 38% of 12th graders nationally scored at or above 302 in reading (May 2014 Nation's Report Card).

For mathematics, Dr. Rosenberg reported the following:

“Given the design, content, and characteristics of the NAEP 12th grade mathematics assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness,

- The percentage of students scoring at or above a score of 163 on grade 12 NAEP in mathematics is a plausible estimate of the percentage of students who possess the knowledge, skills, and abilities in mathematics, making them academically prepared for college.
- A score of 163 is between the cut scores for the basic and proficient achievement levels in 12th grade mathematics.”

¹ The *Technical Panel on 12th Grade Preparedness Research Final Report* states that preparedness focuses on academic qualifications measured by NAEP. Readiness includes behavioral aspects of student performance—time management, persistence, and interpersonal skills, for example—that NAEP does not measure. The Governing Board's deliberations have led to a focus on preparedness as a subset of readiness.

Similar to reading results, Dr. Rosenberg reported that in 2013, 39% of 12th graders nationally scored at or above 163 in mathematics (May 2014 Nation's Report Card).

Other features of the college preparedness portion of Dr. Rosenberg's presentation included information about reporting grade 12 results using academic preparedness research findings. Dr. Rosenberg discussed how the validity argument was developed to support proposed inferences for academic preparedness for college. There is an emphasis on the preliminary nature of inferences for reporting purposes. They are not intended to be preparedness standards, and there is a need for continued research to inform reporting of the 2015 results.

Academic Preparedness Definition for Job Training

In job training, the Governing Board commissioned analyses involving how the NAEP reading and mathematics assessments might be used as indicators of preparedness for training for occupations. Dr. Rosenberg noted that the Governing Board established the following working definition for job training: "The knowledge and skills to qualify for placement into job training without need for remediation in math or reading." The research assumed significant training beyond high school but not a bachelor's degree. The research was not intended to address success or completion of job training. There was also no assumption that academic skills for college and job training are the same.

Compared to the number of studies conducted to investigate NAEP as an indicator of academic preparedness for college, various feasibility issues have resulted in far fewer studies investigating NAEP as an indicator of academic preparedness for job training. Further, the findings from the few studies that have been conducted have been less conclusive. First, a content alignment study between 12th grade NAEP and WorkKeys (used to assess job-related skills) found some similarities between NAEP and WorkKeys content but also identified significant differences in both focus and rigor (ACT, 2010a; 2010b). Specifically, the findings indicated that NAEP is broader both in focus and rigor. Also, the WorkKeys exams in Applied Mathematics and in Reading for Information focus exclusively on the application of academic skills in the workplace, whereas NAEP covers a considerably wider range of mathematical topics and reading skills. Second, findings from the Judgmental Standard Setting Study did not produce supportable conclusions about where to set reference points on the NAEP scale to denote the minimum academic knowledge and skills needed for academic preparedness for entering job training for the targeted occupations. Rather, there was significant variability in the cut scores set by replicate panels within and across occupations (Loomis, 2012; WestEd & Measured Progress, 2011; 2012). Finally, findings from a Course Content Analysis Study revealed that the reading and mathematics content identified in the job-training course artifacts is largely included in the grade 12 NAEP frameworks, but the content of the NAEP frameworks is much larger and broader (similar to what was found in the content alignment study between NAEP and WorkKeys). The course artifacts revealed that few NAEP objectives are covered in the job-training programs (WestEd & Educational Policy Improvement Center, 2013).

Key Takeaways: Academic Preparedness for Job Training

Dr. Rosenberg noted that takeaways based on the content alignment, standard-setting, and course content analysis studies suggested the following:

- Prerequisite knowledge, skills, and abilities (KSAs) for academic preparedness for job training are largely included in NAEP frameworks, but NAEP frameworks are much larger and broader.

- Most of these prerequisite KSAs are measured by NAEP but represent a varying portion of the objectives.

There is no evidence that academic preparedness for job training equals academic preparedness for college. However, students prepared for college are likely to be academically prepared for these job-training programs. Designing studies for academic preparedness for job training is especially challenging.

Dr. Rosenberg concluded that the next steps for job training should consider findings and recommendations from the Framework Evaluation Project. She also recommended the following question for current consideration: Should the Governing Board continue to perform research on using NAEP as a measure of academic preparedness for job-training programs?

Study Update: Comparison of NAEP and O*NET Knowledge, Skills, and Abilities (KSAs)

Dr. Sinclair provided an overview of a second task within the Framework Evaluation Project (Sinclair, Becker, McCloy, & Thacker, 2014). The purpose of the *Comparison between NAEP and the O*NET on Academic Preparedness for Job Training* study was to identify relevant linkages between (a) NAEP reading and mathematics content and (b) training performance required for five target occupations (Automotive Master Technicians; Computer Support Specialists; Heating, Ventilation and Air Conditioning Technicians (HVAC); Licensed Practical Nurses (LPNs); and Pharmacy Technicians). In addition, the study sought to compare the levels of knowledge, skills, and abilities (KSAs) required for the relevant NAEP reading and mathematics content to the levels of KSAs required for the job training content. The KSAs included in the study came from the Occupational Information Network (O*NET), which is the U.S. Department of Labor's occupational information database.

For this study, experts from each target occupation reviewed that occupation's task list from O*NET for its appropriateness to job training. This review was necessary because the O*NET tasks describe *job* performance requirements rather than *training* performance requirements, and the focus of the Governing Board's research is preparedness for job training. Based on the feedback from the occupational experts, the O*NET task lists were edited to ensure their applicability to job training. Next, occupational experts used these lists to identify NAEP reading and mathematics content that is relevant ("linked") to training performance requirements. The occupational experts also identified the training performance requirements that are relevant ("linked") to NAEP reading and mathematics content. Irrelevant content was removed from further consideration. Finally, trained project analysts used academically relevant KSAs from O*NET to systematically rate the levels of KSAs needed for the relevant NAEP reading and mathematics content and the levels of KSAs needed for the relevant job training content. Disconnects between the levels of KSAs needed for NAEP reading and mathematics and the levels needed for job training were flagged for discussion.

The study offered the following findings:

- NAEP-required reading and math skills (grade 8 and grade 12) are broader than skills required for job training.
- NAEP reading objectives most relevant to job training content are those associated with the "locate/recall" cognitive target.
- NAEP reading objectives least relevant to job training content are those associated with the "critique/evaluate" cognitive target.

- NAEP math objectives most relevant to job training content are those associated with the “Number Properties and Operations” (except for computer support specialists) content area – this is true for both grade 8 and 12 NAEP.
- NAEP math objectives least relevant to job training content are those associated with geometry (except for HVAC) and algebra (except for LPNs) – this is true for both grade 8 and 12 NAEP.
- The percentage of the math objectives linked to occupations decreased considerably from grade 8 to grade 12, indicating that as the complexity of the objectives increased from grade 8 to grade 12 their relevance to job training decreased.
- The largest disconnects were found between grade 12 NAEP math and job training. Disconnects also occurred between grade 12 reading and job training. The disconnects in required levels of KSAs tended to be smaller when comparing grade 8 content to job training content, particularly for grade 8 reading.

Dr. Sinclair suggested that the findings call into question the validity of inferences that can be made about using NAEP to report on the preparedness of U.S. 12th grade students for entry into job training. She noted that 12th grade NAEP does not support inferences for using it as an indicator for job preparedness. However, converging evidence across studies shows that the “Number Sense” and “Operations and Measurement” subtopics for math, and the “locate/recall” cognitive target from NAEP informational reading, are most relevant to job training. Therefore, it might be worth considering the use of subscores from these content areas to report on students’ academic preparedness for job training. Specific research studies could be designed to investigate the links between these subscores and initial performance in job training programs.

Study Update: Comparison of NAEP and WorkKeys

Dr. Thacker described a third task within the Framework Evaluation Project – a parallel alignment study that compared NAEP grade 8 and 12 assessments to WorkKeys. WorkKeys is ACT’s job skills assessment that businesses use to measure workplace skills of employees and job applicants, and that schools use to help prepare students for the workplace. The alignment study was designed to determine the extent to which NAEP and WorkKeys measured similar content, and represented an expansion of the initial alignment study with WorkKeys (ACT, 2010a; 2010b) by including NAEP grade 8 and additional content areas from WorkKeys. Overall, the study found similar content between NAEP and WorkKeys but found significant differences in both rigor and focus. NAEP was judged to be considerably broader in content and more academically focused. Although the content areas measured by WorkKeys are largely included in the NAEP framework, the WorkKeys assessments are narrow in terms of which parts of the framework they assess. Adding grade 8 NAEP did not appreciably alter the overall findings from the earlier alignment studies.

The technical report for the alignment study was submitted on November 21, 2014.

SUMMARIES OF PAPER DISCUSSIONS

Drs. Plake, Ryan, and Tippins presented research ideas intended to address ways in which the Governing Board might research issues related to using NAEP's 12th grade results as an indicator of preparedness for entry into career training. The authors, TAP, Governing Board staff, and HumRRO staff engaged in lively discussion of the various ideas related to each paper. The summaries that follow are organized around the major themes of each paper.

Using 8th and 12th Grade NAEP to Measure Student Readiness for Careers ***Barbara Plake, University of Nebraska – Lincoln***

Various research studies have attempted to answer whether high school graduates are college- and career-ready. Dr. Plake proposed a method to determine how well performance on 8th and 12th grade NAEP can support inferences about student job readiness.²

Dr. Plake briefly described the Judgmental Standard Setting Study (WestEd & Measured Progress, 2011; 2012) and the Course Content Analysis Study (WestEd & Educational Policy Improvement Center, 2013), as a preface to her research idea. Both studies focused on the five target occupations identified for NAEP job training preparedness research: Automotive Master Technician, Computer Support Specialist, HVAC, LPN, and Pharmacy Technician. For the JSS study, a judgmental (bookmark) standard-setting approach was employed to set cut scores reflective of career readiness. NAEP items were used to create two parallel forms, arranged in order of increasing difficulty. Then, experts inserted a bookmark where the minimum prerequisite skill level for the job-training program was estimated to be reached (see Appendix D, slide 4). For the Course Content Analysis Study, the goal was to develop descriptors of what students need to know and be able to do overall, based on course and training materials. From both studies there appeared to be a lack of alignment between the skills needed to be successful in selected job-training programs and those measured by the 12th grade NAEP assessment.

Using the existing work as a foundation, Dr. Plake proposed an alternative method to the JSS involving scale anchoring. The goal of this approach is to synthesize the skills that map to career readiness that are characteristic of grades 8 and 12 mathematics and reading items within each of the three NAEP performance categories – basic, proficient, and advanced.

Dr. Plake noted the scale-anchoring method would use the three-parameter logistic (3PL) model for multiple-choice items and a Generalized Partial-Credit Model (GPCM) for constructed-response items. She noted how items are classified into performance categories based on their difficulty parameters:

- *Multiple Choice*: Whichever performance category the item falls into using its difficulty parameter is considered its performance-level classification.
- *Constructed Response*: Each possible score receives a scale and location parameter; the location parameter is used to place the score in one of the score scale latent trait intervals (e.g., performance categories).

Once all items and score values are identified within the performance categories, items would be analyzed by content area. Dr. Plake suggested using the same five job-training programs as used in the prior job preparedness studies. Should the study be successful, the outcome would

² Preparedness for job training. Dr. Plake used the term “readiness” throughout her presentation.

be a better understanding of how, and whether, it is reasonable to use 12th grade NAEP as a vehicle to measure career readiness.

Following Dr. Plake's presentation, there was considerable discussion surrounding the proposed scale anchoring method. Although the TAP did not strive to reach consensus, several considerations of the proposed study were offered, including the following:

1. Considering a matrix arrangement

It was suggested that the domains of the study are broad. Another approach might be to analyze subscores or consider a matrix arrangement where items are mapped onto categories other than the NAEP performance levels. Judges could then rate the importance of the items.

A useful extension might be to use scale anchoring and determine how different strands cluster within each content area. For example, once items are clustered into the major performance level categories, they can be divided into sub-categories and analyzed by strand.

2. Using achievement level categories

Categorizing items by performance level rather than using scale scores follows the basic approach used in the JSS Study. The advantage of this approach is that rather than ordering items and inserting bookmarks at the cut point of proficient, it inserts thresholds that distinguish basic, proficient, and advanced performance levels. By clustering items into larger "buckets," it increases the precision of item difficulty (*b*) parameter estimates. It does not lend itself to over interpreting the location of the parameters.

3. Achieving unidimensionality of the scale

There is considerable concern over the unidimensionality assumption of the scale. That is, there are potential issues analyzing data and reporting results that cross 8th and 12th grades. Dr. Wise suggested that rather than inventing a new standard setting method, it seems most practical to conduct the research at the strand level. Dr. Hambleton stated that it is indeed more likely that the unidimensionality assumption will hold at the strand level. It would also be informative because it will show which KSAs are articulated from the NAEP assessment when looking across grades 8 and 12. It will also allow items to be mapped back to KSAs for the job-training programs.

Grit: A Useful Concept in College and Career Preparedness?
Ann Marie Ryan, Michigan State University

For more information on Grit, see [Angela Lee Duckworth's TED talk](#).

Dr. Ryan defined *grit* as "...perseverance and passion for long-term goals" (Duckworth & Quinn, 2009, p. 166) despite adversity, and stated that it entails "...working strenuously toward challenges, maintaining effort and interest over years despite failure, adversity, and plateaus in progress" (Duckworth et al., 2007, p. 1088). This concept is related to the Governing Board's goals such that often high school graduates are "prepared" and go on to enter college or job-training programs but do not always continue through to completion of that program. Recent research has been exploring the non-cognitive components associated with psychological and mental preparedness, one of which might be *grit*. When assessing the preparedness of 12th graders for college and job-training programs, characteristics related to "perseverance" and "resilience" or grit may have value.

The presentation was divided into the following sections. The summaries that follow are organized around each section.

Conceptual Description

Dr. Ryan observed that measures of grit have two factors – consistency of interest and perseverance of effort. The correlation between these factors has been reported to be .59, yet the few research studies available rarely address the two factors, using an overall *grit* scale instead.

The emphasis of *grit* is sticking with some goal, usually a long-term goal (e.g., college, training program). Characteristics of gritty people include a willingness to engage in greater practice and a capacity to continue with repetitive or boring tasks. These people are also more likely to persevere in the face of negative feedback or setbacks.

Prototypical Measurement

Dr. Ryan noted that the most reported instrument used to measure grit is an 8-item self-report scale, the Grit-S. This measure has an overall internal consistency reliability of .70 or higher. The Grit-S was developed by selecting items from four samples that had used the Grit-O (a 12-item form), looking at item-level correlations with outcomes (retention, GPA), and then eliminating two items from each of the two subscales that were most frequently below the median in predicting outcomes across the four samples. In addition to the Grit-O, the Grit Grid, resume rubric, and interview are all used to measure grit.

Although interviews have been used to measure *grit*, Duckworth suggested the interview is a poor measure because a gritty person would not likely self-report they are gritty. (Of course, this also calls into question the viability of self-report measures.) There has not been any evidence of vulnerability to socially desirable responding.

Construct Distinctiveness Evidence

Empirical evidence has supported the idea that *grit* is a facet of a well-researched personality construct, conscientiousness. Six facets of conscientiousness and grit scales are highly correlated. Studies have also researched whether grit was related to

- Industriousness (e.g., working hard or being organized) – Data are inconclusive.
- Self-control – Grit is a higher-level goal, whereas self-control is immediate (e.g., should I eat this cookie or not).
- Cognitive ability – There is little evidence of a relation in the literature.

However, Dr. Ryan also noted that considerable research efforts have focused on whether grit is something other than conscientiousness. Duckworth and others have argued grit is different because individuals can be conscientious in the short-term but not stick with things over time. In addition, incremental validity evidence suggests that grit is something more than conscientiousness, given it has been shown to provide incremental prediction to outcomes when conscientiousness is included.

Connection to Outcomes: Performance and Attrition

Dr. Ryan's discussion of grit and accomplishments focused on educational attainment, GPA, teacher effectiveness, and advancement in competition. Additionally, she highlighted factors relating to grit and persistence behaviors (i.e., hours of practice and persistence at a task). She noted how grit may predict such outcomes as health and well-being, burnout, and suicide ideation. Most of the studies linking grit to outcomes focus on either retention or accomplishment (e.g., academic program completion; career changes, divorce), with relations in expected directions (e.g., positively related to completion; negatively related to career change, divorce [for men], and attrition).

Group Differences

Studies have researched grit and group differences and found older people are grittier than younger people. One might speculate this is due to older people knowing what it takes to persist in tasks.

Relation to Academic and Workplace Training Preparedness

It appears evidence for grit is growing, with more studies showing it relates to retention and accomplishment outcomes. Dr. Ryan noted that additional research should be focused in areas such as

- Broad v. facet use in prediction
- Domain specific v. domain general – Is it possible to have grit in one domain but not in others?
- Self-selected v. assigned goals – How predictive is grit under these conditions? For example, high school does not include many self-selected goals.
- Feedback seeking and support seeking
- Interventions – If grit is a malleable trait, how could we intervene?
- Negative effects of grit – Does grit depict a curvilinear relation with outcomes (e.g., an inverted U function)?

Concluding her presentation, Dr. Ryan turned the focus back to the initial research question: How useful is grit in terms of understanding the preparedness of 12th graders for college and job-training programs? Dr. Ryan proposed three considerations. First, grit is one exemplar of a motivational trait that better explains how and when some 12th graders persist in college or training programs. Second, in college or postsecondary settings, students have more discretion

in self-selected goals (e.g., courses or training) than they do in high school; as a result, grit might be of value when considering the goals students choose to engage in. Third, grit fits into the Partnership for the 21st Century framework (P21; for more information, visit [the P21 website](#)). Specifically, grit is relevant to Life and Career skills, which include managing goals and time, working independently, and being a self-directed learner.

There was considerable discussion surrounding grit and how it relates to preparedness for college and job training programs. Several topics were discussed, including the following:

1. Grit and OCD behaviors

The question was raised whether there is anything in the literature about the relation between grit and behaviors associated with obsessive-compulsive disorder (OCD). It was hypothesized that a curvilinear relationship might exist between the two constructs in that excessive grit might be associated with OCD behaviors.

2. Persistence and grit

Research does not suggest people with grit will succeed; it suggests only that people will persist. Some individuals will persist in light of negative feedback, but it might be valuable to examine the amount or types of negative feedback individuals who “persist” receive. It should be considered that, in some instances, it may be detrimental to encourage someone who may be gritty but lacks basic skills to pursue some goal.

3. Alternate methods for measuring grit

Self-reported measures might not be the best approach. Individuals are more likely to *fake* motivation for grit on employment/job skill measures. Could grit measures be included with other ability measures?

4. Grit and components of the Motivation performance determinant

Dr. McCloy noted that the three dimensions of grit identified in the presentation align well with the three components of the motivation performance determinant in the Campbell, McCloy, Oppler, and Sager (1993) model of job performance: choice to expend effort, choice of level of effort to expend, and choice to persist in the expenditure of that level of effort. To the extent this is true, perhaps grit is not so much a personality trait but rather a characteristic more amenable to improvement/training

5. Implications of including grit as component of academic preparedness

If grit is considered a contributing factor to academic preparedness, then the question was raised as to how much weight should be given to grit in determining academic preparedness. Dr. Ryan noted research showing that grit correlates more highly with outcomes when individuals have chosen to engage in the activity. Consequently, grit should be of greater value (i.e., given more weight) for activities individuals have chosen. High-school-level work, including taking 12th grade NAEP, does not generally constitute activities that individuals have self-selected. Consequently, this should be kept in mind with regard to the question of weighting.

Relating NAEP to Commercial Off-the-Shelf Measures
Nancy Tippins, Corporate Executive Board – Valtera Corporation

Past research on NAEP suggests that some test items are not related to success in job training and concludes that the NAEP framework for math and reading is broader than that required for many job-training courses. At the same time, there are likely other knowledge, skills, abilities, and other characteristics (KSAOs) that are related to preparedness in specific fields and jobs. In particular, “soft skills” such as perseverance, determination, conscientiousness, motivation for continuous learning, service orientation, safety awareness, and teamwork have not been factored into assessments of preparation for training.

In her proposed research study, Dr. Tippins explored relations between NAEP test scores and preparedness for job training courses in business and industry. The approach consists of data analysis and one administration of NAEP. The questions this study would address include the following:

- To what extent are NAEP scores related to measures of job-preparedness?
- To what extent are NAEP scores related to success in job training and/or certification?
- Are the levels of reading and math items in NAEP consistent with the levels necessary for success in job training and/or certification?
- What are the KSAOs (knowledge, skills, abilities, and other characteristics) necessary for success in training that are not measured by NAEP?

The process to execute this study would rely on a job analysis and statistical analysis. Several job preparedness measures would be collected at three points in time: prior to employment, during training, and after training.

Dr. Tippins explained that a great deal of effort is required to validate and implement operational pre-employment tests, create effective training programs, and develop certification programs. Therefore, it would be costly to take this approach. This research would rely instead on the use of an organization’s existing certification program. To be feasible, the organization’s program should meet certain requirements. Dr. Tippins recommended selecting an organization that has done some of the following:

- conducted a job analysis that identifies the KSAOs required to perform the job on day one (i.e., prior to training),
- ensured that the level of math and reading is evaluated in the job analysis,
- administered pre-employment tests and systematically retained the test data,
- offered multiple-choice tests,
- offered performance tests,
- administered certification tests and systematically retained test data,
- provided supervisor ratings of performance, and
- developed objective key performance indicators (KPIs).

Dr. Tippins suggested that once an organization was identified, the Governing Board would conduct analyses of the job analysis data. Specifically, SMEs would be responsible for making similarity ratings between NAEP constructs and the KSAOs identified in the job analysis. This would allow the Governing Board to identify gaps between NAEP and job performance. A final

step would be to statistically relate NAEP scores to pre-employment test scores, certification tests, and other associated measures used by the organization.

Finally, Dr. Tippins noted that one of the greatest challenges to this study would be finding an organization to participate. The study must be designed in such a way that there is some benefit to an organization for participation. Things to be considered include (a) costs, particularly with respect to employees' time away from the job to participate in the administrations; (b) confidentiality concerns; (c) unintended consequences – the study might adversely affect the organization, for example, finding gender differences in the performance level of their employees; and (d) replicability across jobs and organizations.

Additionally, Dr. Tippins posed potential concerns involving sufficient measures including inconsistencies in data storage; retests; validity and reliability of employment tests; validity and reliability of certification tests; and regional differences, noting a need to standardize KPIs data, restriction of range, and unmeasured variables.

Following Dr. Tippins' presentation, there was considerable discussion on exploring the relations between NAEP test scores and preparedness for job training courses. Several ideas were considered, including the following:

1. Identifying organizations willing to participate

Dr. Tippins noted that organizations that do pre-employment testing for all of their entry-level jobs often use commercial, off-the-shelf tests. She suggested the Governing Board might identify willing organizations by going to test publishers and finding out who their clients are and who uses their tests. The Governing Board will want to identify an organization that uses pre-employment testing for all entry-level jobs.

2. Item selection

In light of past NAEP research showing that some test items are not related to success in job training and suggesting that the NAEP framework for math and reading is broader than that required for many job-training courses, the question was raised whether there is merit in handpicking items out of a pool or picking blocks that seem representative of NAEP content. One approach to evaluating the extent of the relation would be to compare the constructs measured by NAEP with the KSAOs required for training success. This approach relies on careful job analysis of entry-level jobs to identify the KSAOs required by the jobs and a comparison of them to the constructs NAEP measures. An important outcome of this part of the study is the identification of KSAOs that are important to training and job success that are not measured by NAEP. If NAEP is broader and includes content that does not relate to job training, those items need not be administered.

3. Benefits to the industry or participating organizations

One reason an organization might choose to participate in this study is for the good of the country and the educational system. Communicating the idea that what is learned in schools is also relevant to what you need in the workforce would present a clear path for students to be able to say, "I aspire to be a whatever (e.g., cable technician), so I'll need to focus my studies on XYZ." Unfortunately, there is no bottom-line or direct benefit to the organization. (Note that other benefits could be created during the study, however, such as agreeing

to analyze the organization's data beyond the level that they would be able to produce on their own.) Dr. Tippins recommended that to improve chances of participation, it would be important to be flexible—for example, consider allowing employees to participate at a time that is convenient for them (i.e., do not designate a specific day or time); arrange for employers to have a considerable timeframe for allowing employees to complete the test (e.g., a year); and consider administering the test over the internet. Dr. Crovo noted that NAEP is transitioning to technology-based platforms for math in 2017–2019, which might make this a possibility.

GROUP DISCUSSION

After the paper presentations, Dr. McCloy opened the symposium up for a group discussion for the Governing Board and TAP members to comment further on the paper presentations or on any technical issues for NAEP preparedness research. This summary highlights the major themes reflected during the discussion.

Crossing students/tests with employment settings/tests to establish standards

Dr. Gong asked about the purpose of an empirical study relating the NAEP measures to job performance, selection, and certification. He noted that the NAEP and a set of employment tests should be administered to a common sample of students. It was recommended that consideration be given to programs that have information about qualification cut scores. Rather than administering tests to collect data, the focus should be how colleges use ACT scores to determine class placement.

Reporting out on preparedness

Dr. Crovo stated that it would be desirable to be able to report on preparedness if we could demonstrate that NAEP is a reliable and valid measure of preparedness. This would help us attain the Governing Board's overall goal, which is to improve the usefulness/utility of NAEP. The primary audience for NAEP is the general public, so the question becomes what can the Governing Board do to make results more meaningful and useful to the general public? Preparedness might be a key to attaining that goal.

Reporting results (individual level vs. aggregate level)

Dr. Gong suggested that it is better to discuss aggregate results (state level) than to articulate results for individuals. There may be little value in reporting results at an individual level. Dr. Crovo and Dr. Rosenberg explained that during NAEP administrations, before test proctors leave the school with data, student identification information is removed and left at the test site (school). Student data are not kept with the overall NAEP results. Consequently, there is no way to track students into the future without special permission (as has been obtained in Florida). Dr. Crovo noted, however, that ideally we would be able to report on subgroup differences.

Policy issues concerning sharing student data

Dr. Gong suggested going through state agencies as an alternative to data sharing agreements. However, this raises the concern that states may question why we would want to do this. One suggestion was to show states how higher NAEP scores tend to correspond with other indicators of state-level success (e.g., economic growth, median income, unemployment rates). Dr. Thacker suggested using longitudinal studies to demonstrate that states that score higher and/or improve on NAEP tend to improve on these other indicators as well.

Subgroup performance

Panel members discussed that it might be beneficial to determine the aggregate level of subgroup performance in comparison to grit or another measure that might serve as a predictor. Dr. Crovo suggested looking at subgroup relations with these measures (e.g., selection or promotion assessments in large corporations) and performance on various NAEP subjects. Dr. Thacker suggested looking at subgroups such as socioeconomic status (SES), gender, and ethnicity. Dr. Tippins interjected that cooperation from corporations would likely decline if we opted to use SES because such data are usually not collected and stored.

Indicators of success

Dr. Gong changed the direction of the discussion to focus on indicators of success, suggesting the Governing Board might want to consider indicators, and analysis of achievement levels, linking them to occupations. Some indicators could be anchored by occupation, with state reports displaying the average number of students who applied for each occupation. Dr. Crovo cautioned that we do not want to say that to be prepared, you must first be proficient, or that a particular occupation accepts students who do not score at the NAEP proficient level. Dr. Gong noted that WorkKeys is designed to yield a score that shows you are qualified for particular occupations. If we applied the same process to NAEP skills scores (perhaps using O*NET), we might be able to achieve something similar. Implicit in this process is the assumption that preparedness for job training would be occupation-specific.

NEXT STEPS

During this symposium, many challenges were discussed with regard to conducting research on academic preparedness for job training. For example, the variability in training programs across occupations—and even the variability across training programs within the same occupation—makes it exceedingly difficult to determine a reference point on the NAEP scale that signifies academic preparedness for job training. Furthermore, the fact that students do not take all items that represent any given NAEP construct, combined with the lack of student-level scores from NAEP, adds further challenge to conducting a criterion-related validation study where individuals' NAEP scores are correlated with some measure of job training. Even if special arrangements could be made to administer a special NAEP test representative of the full content domain and for which individual-level scores could be produced, it would likely be difficult to obtain participation from organizations to support this special administration. It also would be challenging to obtain scores from the NAEP respondents on some validated measure of job training. Challenges such as these, coupled with research findings that have provided converging evidence that NAEP measures reading and math content that is larger and broader than the reading and math KSAs required for entry into job training, call into question whether the Governing Board should continue to conduct research on using NAEP as a measure of academic preparedness for job-training programs.

If, however, the Governing Board decides to move forward with preparedness research for job training, the next agenda items to consider based on the TAP symposium would be the following:

1. Build upon Dr. Plake’s idea to use NAEP to measure student job readiness

Dr. Plake suggested a modification of prior expert judgment research—namely, asking job training experts to evaluate the importance for training of items mapped onto the NAEP Achievement Levels, rather than simply rating discrete items as was done previously. The items might be organized by content strand, so that importance might be rated differentially by strand. The general idea is to get experts to respond to more general descriptions of academic preparedness, rather than to individual items.

2. Collect information on additional aspects of preparedness through student (and teacher?) questionnaires

Further research on academic preparedness for job training might employ measures of the grit construct (see Dr. Ryan’s presentation) as potential moderators.

3. Analyze/conduct longitudinal research relating academic proficiencies at the end of high school to subsequent career success

The National Center for Education Statistics (NCES) has conducted several research studies following students from high school into post-high school activities. The National Longitudinal Surveys conducted by the Department of Labor contain similar data. Results from prior analyses of these data, together with new analyses, could inform the relation between academic skills in high school and persistence and success in different career fields.

4. Consider studies suggested by Dr. Tippins linking employment tests to the NAEP scale

The approach would be something along the lines of the following: (a) recruit a few large companies that provide entry-level jobs not requiring prior job knowledge or experience, (b) ask them to administer a number of NAEP items along with commonly used employment tests so that the employment test and NAEP scales can be linked, (c) collect information on key decision points, and (d) summarize findings on the relation between the employment test score levels and subsequent success.

5. Embed preparedness items into NAEP

Dr. Champion recommended embedding items from a measure of job training preparedness (e.g., WorkKeys) into the NAEP assessment. This could offer an alternative to trying to get job trainees and/or individuals in associate degree programs to complete a special administration of NAEP.

Dr. McCloy and Dr. Crovo closed the symposium by thanking panel members, Governing Board staff, and HumRRO staff for attending and contributing to the meeting. The discussions yielded a lot of valuable information and insightful perspectives to propel the Governing Board forward in preparedness research.

REFERENCES

- ACT, Inc. (2010a). *The alignment of the NAEP grade 12 mathematics assessment and the WorkKeys applied mathematics assessment*. Iowa City, IA: Author.
- ACT, Inc. (2010b). *The alignment of the NAEP grade 12 reading assessment and the WorkKeys reading for information assessment*. Iowa City, IA: Author.
- Campbell, J.P., McCloy, R.A., Oppler, S.H., & Sager, C.E. (1993). A theory of performance. In N. Schmitt & W. Borman (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco, CA: Jossey-Bass.
- Loomis, S. (2012, April). *A study of "irrelevant" items: Impact on bookmark placement and implications for college and career readiness*. Presentation at the annual meeting of the National Council on Measurement in Education, Vancouver, Canada.
- Sinclair, A.L., Becker, S., McCloy, R.A., & Thacker, A.A. (2014). *Comparisons between NAEP and O*NET on academic preparedness for job training for five target occupations*. (HumRRO Report 2014 No. 12). Alexandria, VA: Human Resources Research Organization.
- WestEd & Measured Progress. (2011). *National Assessment of Educational Progress Grade 12 Preparedness Research Project Judgmental Standard Setting (JSS) Studies: Process Report*. San Francisco, CA: Author.
- WestEd & Measured Progress. (2012). *National Assessment of Educational Progress Grade 12 Preparedness Research Project Judgmental Standard Setting (JSS) Studies: Technical Report*. San Francisco, CA: Author.
- WestEd & Educational Policy Improvement Center (2013). *National Assessment of Educational Progress Grade 12 Preparedness Research Project Job Training Programs Curriculum Study*. Eugene, OR: Author.

APPENDIX A

MEETING AGENDA

**NAEP 12th Grade Reading and Mathematics Framework Evaluation Project
Technical Advisory Panel (TAP) Symposium
August 20, 2014
Fairfax at Embassy Row
Washington, DC**

8:30 am-9:00 am	Registration/Coffee
9:00 am-9:15 am	<i>Welcome and Introductions</i> Rod McCloy, HumRRO; Mary Crovo, NAGB
9:15 am-9:45 am	<i>Update on NAEP 12th Grade Preparedness Research</i> Sharyn Rosenberg, NAGB
9:45 am-10:00 am	<i>Overview of NAEP Framework Evaluation Project</i> Andrea Sinclair & Art Thacker, HumRRO
10:00 am-11:00 am	PAPER: <i>Using 8th and 12th Grade NAEP to Measure Student Readiness for Careers</i> Barbara Plake, University of Nebraska—Lincoln, retired, private consultant
11:00 am-11:15 am	BREAK
11:15 am-12:15 pm	PAPER: <i>Grit: A Useful Concept in College and Career Preparedness?</i> Ann Marie Ryan, Michigan State University
12:15 pm-1:15 pm	WORKING LUNCH <i>Continued Discussion of Preparedness Research</i>
1:15 pm-2:15 pm	PAPER: <i>Relating NAEP to Commercial Off-the-Shelf Measures</i> Nancy Tippins, Corporate Executive Board – Valtera Corporation
2:15 pm-2:30 pm	BREAK
2:30 pm-3:15 pm	<i>Group Discussion: (1) Paper presentations and (2) Technical issues for NAEP preparedness research</i>
3:15 pm-3:30 pm	<i>Closing Remarks</i> Mary Crovo, NAGB; Rod McCloy, HumRRO


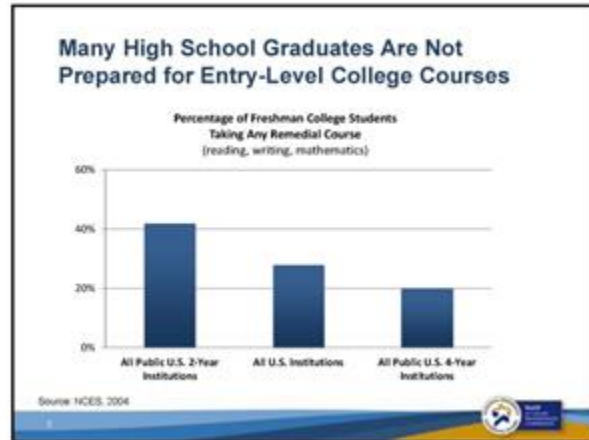
APPENDIX B

UPDATE ON NAEP AND ACADEMIC PREPAREDNESS FOR COLLEGE AND JOB TRAINING (PRESENTATION)

Update on NAEP and Academic Preparedness for College and Job Training

Presentation to the Technical Advisory Panel
August 20, 2014

Sharyn Rosenberg, Ph.D.
Assistant Director for Psychometrics


NAEP and Academic Preparedness

Are the nation's 12th graders well-prepared academically for college and job training?

The U.S. has no common definition or measure of 12th grade preparedness

↓

Governing Board program of research:
Transform 12th grade NAEP into indicator of preparedness for postsecondary education and training




The Governing Board's Preparedness Research

Stipulations

- 4 Postsecondary education and training in the U.S. is a complex enterprise of independent institutions and organizations, public, private, and proprietary.
- 4 Across and within institutions, different admission requirements and different policies used for placing individual students into courses and majors.
- 4 There is no single, generally accepted definition of "academic preparedness for college" or for "job training" in the U.S.
- 4 Therefore, a working definition was needed to design/conduct the NAEP research.
- 4 *Did not assume:* Academic skills for college and for job training necessarily the same.
- 4 *Did assume:* Education systems intend all H.S. students to graduate able to pursue the path they choose, whether college or job training.



A Comprehensive Plan: More Than 30 Studies in Five Research Areas

Study Types	Questions
Content Comparison	Is the content of NAEP similar to relevant tests?
Statistical Relationships	How does performance on NAEP compare to performance on other relevant tests?
Judgmental Standard Setting	What is the point on the NAEP scale that experts judge as just "academically prepared"?
Survey of Postsecondary Institutions	What are the tests and cut-scores used for placement?
Benchmarking	How do selected reference groups perform on NAEP?



College



Working Definition: Academic Preparedness for College

The knowledge and skills to **qualify for placement** into entry-level, credit-bearing college courses **without need for remediation** in math or reading

- **"Preparedness"** versus **"Readiness"** – NAEP measures academic proficiency
- Qualify for placement
- No assumption of success or completion
- Placement vs. admissions
- "Typical" student in "typical" college



Content Comparisons: College

ACT, SAT, ACCUPLACER

- NAEP content is similar to the ACT, SAT, and ACCUPLACER.
- ACT, SAT, and ACCUPLACER content is generally covered by NAEP.
- NAEP is generally broader.
- NAEP differs in some ways:
 - types of test questions
 - cognitive rigor of test questions
 - types of reading texts
 - weighting and coverage of specific content in mathematics



Statistical Comparisons: College

NAEP/SAT 2009

- NAEP and SAT mathematics = **.91** correlation
- NAEP and SAT reading = **.74** correlation

NAEP High School Transcript Study 2005/2009

- NAEP to ACT and SAT results confirm NAEP/SAT linking study

Florida longitudinal study

- First year college GPA
- Placement into remedial/developmental
- ACT and SAT college readiness benchmarks



Other Research: College

Course Content Analysis: First Year College Courses

- Review of course syllabi and textbooks for reading and math demands

Benchmarking

- Texas postsecondary pilot study

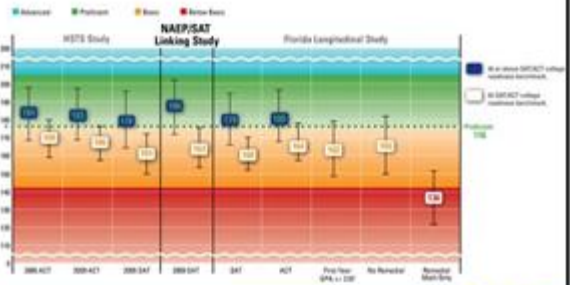
Higher Education Survey

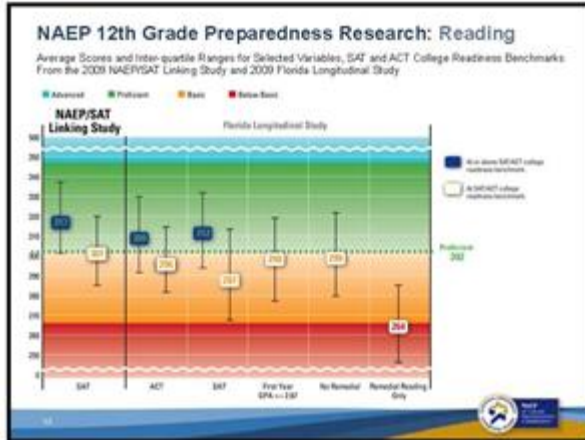
- Tests and cut-scores used to determine need for remedial in college
- Only extant nationally representative survey
- Results show variability in cut scores for "needs remedial"



NAEP 12th Grade Preparedness Research: Mathematics

Average Scores and Inter-quartile Ranges for Selected Variables, SAT and ACT College Readiness Benchmarks From the 2009 NAEP/SAT Linking Study, 2005 High School Transcript Study, 2009 High School Transcript Study, and 2009 Florida Longitudinal Study





Academic Preparedness for College: Key Takeaways

- Content comparisons confirm appropriateness of revised reading and mathematics frameworks
- Students scoring at or above *Proficient* likely to be academically prepared, *but...*
- Results are different for reading and math in relation to *Proficient* and interpretations/implications for reporting need to be considered

Reporting Grade 12 Results Using Academic Preparedness Research Findings

- Validity argument was developed to support proposed inferences for academic preparedness for **college**
 - Emphasis on preliminary nature of inferences for reporting purposes
 - Not intended to be preparedness standards
 - Continued research to inform reporting of 2015 results
- At the August 2013 meeting, Board members voted on a motion to approve the following statements for use in the reporting of the NAEP 12th Grade Reading and Mathematics assessments for 2013:

Reading

- Given the design, content, and characteristics of the NAEP 12th grade reading assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness:

the percentage of students scoring at or above a score of 302 on Grade 12 NAEP in reading is a plausible estimate of the percentage of students who possess the knowledge, skills, and abilities in reading that would make them academically prepared for college.

A score of 302 corresponds to the cut-score for the Proficient achievement level in 12th grade Reading.

In 2013, 38% of 12th graders nationally scored at or above 302 in reading (May 2014 Nations Report Card)

Mathematics

- Given the design, content, and characteristics of the NAEP 12th grade mathematics assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness:

the percentage of students scoring at or above a score of 163 on Grade 12 NAEP in mathematics is a plausible estimate of the percentage of students who possess the knowledge, skills, and abilities in mathematics that would make them academically prepared for college.

A score of 163 is between the cut-scores for the Basic and Proficient achievement levels in 12th grade mathematics.

In 2013, 39% of 12th graders nationally scored at or above 163 in mathematics (May 2014 Nations Report Card)



Next Steps for Academic Preparedness for College: Grade 12 NAEP

Research planned and underway for 2013 - 2015:

- National NAEP-ACT linking study
- State partners at grade 12 (FL, IL, MA, MI, TN)

Current consideration: How should research findings from phase two be used to inform reporting of the 2015 NAEP results in terms of academic preparedness for college?



Next Steps for Academic Preparedness for College: Grade 8 NAEP

Research planned and underway for 2013 - 2015:

- NAEP-Explore content alignment study
- State partners at grade 8 (KY, NC, TN)

Current consideration: How should the findings from research with grade 8 NAEP be used to report on the percent of students "on track" for academic preparedness for college?



Job Training



Academic Preparedness Definition for Job Training

The knowledge and skills to **qualify for placement** into job training **without need for remediation** in math or reading

- **"Preparedness"** versus **"Readiness"** – NAEP measures academic proficiency
- Goal: identify prerequisite KSAs and reference point on NAEP scales
- **"Job training"** versus **"Career"**
- Assumption: significant training beyond high school but not B.A.
- No assumption of success or completion
- No assumption academic skills for college and job training necessarily the same



Content Comparisons: Job Training

WorkKeys

- Some NAEP content is similar to WorkKeys.
- WorkKeys measures some content that NAEP does not.
- NAEP is broader.
- WorkKeys Applied Mathematics and Reading for Information focus on the application of foundational skills in workplace situations and NAEP does not.
- NAEP does not have workplace situations as its primary focus.



Standard Setting: Job Training

Ten studies in five job training programs

Programs included

- Automotive Master Mechanic
- Computer Support Specialist
- HVAC Technician
- Licensed Practical Nurse
- Pharmacy Technician

Criteria for selection

- Significant numbers, stable or growing positions
- Range of occupations, civilian and military comparability
- Training beyond high school, not bachelor's degree
- Range of reading and math skills
- Compensation with growth potential
- Familiar to public



Standard Setting: Job Training

Study Results

- Results not confirmed by the replicate panels/other research results
- Panelists deemed many NAEP 12th grade items not required for determining academic preparedness for their job training programs
- Findings do not support conclusion that academic preparedness for these job training programs is same as for college
- Results suggest a need for closer look at course content requirements



Course Content Analysis: Job Training

Study Results

- Large study analyzing course materials from 122 institutions – 85 courses for mathematics content and 80 courses for reading content
- Prerequisite KSAs are largely included in the grade 12 NAEP frameworks, but NAEP frameworks are much larger and broader
- Most course prerequisites are measured by NAEP, but represent a mostly small, varying portion of the framework objectives
 - Between 64% and 77% of the 130 mathematics objectives **NOT evident** as prerequisite in any course examined within the five occupations.
 - Between 16% and 68% of the 37 reading objectives **NOT evident** as prerequisite in any course examined within the five occupations.



Academic Preparedness for Job Training: Key Takeaways

- Content comparison, standard-setting, and course content studies suggest
 - Prerequisite KSAs are largely included in NAEP frameworks, but NAEP frameworks are much larger and broader.
 - Most prerequisite KSAs are measured by NAEP, but represent a varying portion of the objectives.
- No evidence that academic preparedness for job training = college.
- However, students prepared for college likely to be academically prepared for these job training programs.
- Designing studies for academic preparedness for job training is especially challenging.



Next Steps for Academic Preparedness for Job Training

- Consider findings and recommendations from Framework Evaluation Project

Current consideration: Should the Governing Board continue to perform research on using NAEP as a measure of academic preparedness for job training programs?



View the Report

www.nagb.org/what-we-do/preparedness-research.html



For More Information...

- ★ **National Assessment Governing Board:**
www.nagb.org
- ★ **Preparedness Research and Resources:**
www.nagb.org/commission/researchandresources.html



APPENDIX C

USING 8TH AND 12TH GRADE NAEP TO MEASURE STUDENT READINESS FOR CAREERS (PAPER)

Barbara S. Plake
University of Nebraska-Lincoln, Emeritus
August, 2014

Prepared for National Assessment Governing Board

There have been various attempts to garner information about whether high school graduates are college and career ready. The focus of this study is to propose a methodology to determine how well performance on the 8th and 12th grade NAEP assessment can provide evidence to support inferences about students' readiness for careers. In a previous effort to use 12th grade NAEP for this purpose, a standard-setting type panel attempted to set cut scores reflective of career readiness. That effort had many challenges and was not successful in meeting the intended goals. In another study, skill requirements for selected job-training programs were compared to the skills measured with the 12th grade NAEP (National Assessment Governing Board, 2013). There appeared to be a lack of alignment between the skills needed to be successful in these selected job-training programs and those measured by the 12th grade NAEP assessment. One of the recommendations of that study was to consider the skills measured as well by the grade 8 NAEP assessment.

With this intent in mind, this study describes the use of a scale-anchoring methodology on the 8th and 12th grade NAEP assessments to identify the items that best align with and distinguish between the performance levels of Basic, Proficient, and Advanced performance categories. Once these items have been identified, they would be analyzed to see how well the skills measured by these items are consistent with those needed for career readiness. This methodology is different from the one used in the aforementioned study of the alignment of 12th grade NAEP with the job-training programs, which focused on a Bookmark-type Ordered Item Booklet to identify the cut points for career readiness. A scale anchoring methodology clusters items into performance categories based on their item difficulty calibrations. Items in the same performance category are treated as equivalent. The Ordered-Item Booklet methodology also focuses on item difficulty parameters but sequences them numerically along a latent continuum, giving the impression that these item parameters are measured without error.

This scale-anchoring methodology will apply both to multiple-choice (MC) items and to constructed-response items using an IRT approach. Most scale-anchoring studies in the past have focused only on MC items, leaving out the additional information provided by constructed-response (CR) tasks. The methodology that is being proposed is described in Reshetar et al. (2013) and combines IRT parameter estimates from a 3-PL model for multiple-choice items and a Generalized Partial Credit Model (GPCM) for constructed-response tasks. In general, items are classified into performance categories based on their IRT difficulty (or threshold) parameters. For MC items, each item's difficulty parameter (b) is compared with the scale score latent trait and corresponding intervals delineating the three performance categories (Basic, Proficient, and Advanced). Whichever performance category the item's difficulty parameter falls into is considered the item's performance level classification. Using the GPCM, each CR item's possible score (or rubric value) receives a scale and location parameter. The location parameter is used to place the CR item score in one of the scale score latent trait intervals (or performance categories), resulting in classification of all possible score points on each CR item.

Once items and score values for CR-tasks are identified in the performance categories, these items will be analyzed to synthesize skills that map to career readiness and are characteristic of the items within each performance category.

The outcome of this study will be a better understanding of how, and whether, it is reasonable to use 12th grade NAEP as a vehicle to measure career readiness.

References

National Assessment Governing Board (March 2013). *National Assessment of Educational Progress grade 12 preparedness research project job training programs curriculum study*. Washington, DC: Author.

Reshetar, R., Kaliski, P., Chajewski, M., Marsh, R., Chuah, S. C., Coleman, C., & Plake, B.S. (April, 2013). Performance level descriptor validation studies for four Advanced Placement Program (AP) mixed-format exams. American Educational Research Association, San Francisco, CA.

APPENDIX D

USING 8TH AND 12TH GRADE NAEP TO MEASURE STUDENT READINESS FOR CAREERS (PRESENTATION)

Using 8th & 12th Grade NAEP to Measure Student Readiness for Careers

Barbara S. Plake, University of NE, Emeritus

What does performance on NAEP say about job preparedness?

- Job Training Programs Curriculum Study, March 2013
 - Using course and curriculum training materials for 5 job training fields (Automotive Master Technician, Computer Support Specialist, HVAC, LPN, Pharmacy Technician)
 - Articulated KSA needed for these training programs (not limited to NAEP Framework KSAs)
 - Goal: to develop descriptions of what students need to know and be able to do overall, based on these course materials

Judgmental Standard Setting Study

- WestEd & Measured Progress, 2011
 - Identified KSA needed to be Just Prepared for each job category (Borderline Performance Descriptors)
 - The next step in the Job Curriculum Study was to compare the KSAs yielded by the JCS with those determined using the JSS approach
 - Course-related KSA were generally less rigorous than those from the BPD, especially in mathematics.

Methodology for JSS

- Bookmark Standard Setting
- NAEP items were used to create two parallel 12th grade NAEP assessments, Booklets A & B
- Created OIB comprised of items from Booklets A & B in increasing difficulty, with RP = .67
- NAEP experts inserted bookmark where the minimum prerequisite skills for the job training program was reached.

Alternative Approach

- Scale anchoring methodology
- Use 8th and 12th grade NAEP
- Focus on NAEP performance categories (Basic, Proficient, Advanced)
- Identify items that are best aligned to these categories for Reading and Mathematics, for 8th and 12th grade NAEP items
- Can be used with MCQ and CR tasks

Scale Anchoring Methodology

- 3 PL IRT model for MCQs
- Graded Partial Credit Model for Constructed Response tasks with multiple score points
- Items are classified into performance categories based on their difficulty parameters
 - MCQ: whichever performance category the item falls into using its b parameter is considered its performance level classification
 - CR: each possible scores receives a scale and location parameter; the location parameter is used to place the score in one of the score scale latent trait intervals (e.g., performance categories)

Synthesis of Results from Scale Anchoring

- Once all items and score values are identified with performance categories (for mathematics at grades 8 and 12 and for reading at grades 8 and 12), these items will be analysed (by content area)
- Goal: synthesize skills that map to career readiness that are characteristic of the items within each performance level.
- Could use the same 5 job training programs as used in the Job Curriculum Study

Outcome of Study

- Focus on content areas, but combine items from 8th and 12th grade NAEP
- Could lend more evidence to the question of whether it is reasonable to use NAEP performance as an indicator of job preparedness.

APPENDIX E

GRIT: A USEFUL CONCEPT IN COLLEGE AND CAREER PREPAREDNESS? (PAPER)

Ann Marie Ryan
Michigan State University
August 1, 2014

Prepared for National Assessment Governing Board

Introduction

“Men fail much oftener from want of perseverance than from want of talent.” -William Cobbett (1829).

Major accomplishments – graduations, promotions, lifetime achievement awards -- are often described as taking more than skills or intelligence, but requiring hard work, tenacity or perseverance. Along these lines, *grit* is a concept that has received a great deal of interest in the past decade as a potentially important predictor of success, and in particular of “sticking with it” (i.e., as a negative correlate of attrition). Grit’s popularity is in part due to the visibility of researcher Angela Duckworth, who has a popular TED talk on the topic and who is the recipient of a MacArthur “genius grant” for this work. There are also a number of bestseller books and popular blogs on parenting and childhood education that focus on this topic (Davis, 2014; Tough, 2013).

Why might NAGB be interested in grit? Success in postsecondary education settings as well as in job-training programs certainly requires basic skills such as those currently assessed by the NAEP. However, attrition is an oft-noted concern for educational and training programs, as individuals who have the skills to enter these settings do not always complete. For settings that can be intense (e.g., compressed time frame courses, short probationary periods for on-the-job training) or challenging (e.g., programs with high failure rates), being “prepared” involves psychological or mental preparation as well as having the requisite skill set. In higher education, there has been a recent interest in exploring non-cognitive predictors of academic success and retention (e.g., see Schmitt, 2012 for a summary of one program of research on this topic). Thus, in a quest to best assess the preparedness of 12th graders for job training and higher education settings, a consideration of characteristics beyond skills that relate to “sticking with it” may have value. Grit is one such characteristic, and will be the primary focus of this paper.

This paper is organized as follows. First, as an introduction, the most typical conceptual definition of grit will be discussed. Second, the psychometric properties of the prototypical measure of grit will be briefly summarized. Next, evidence of the construct distinctiveness of grit will be described, as well as information on its correlates. Fourth, evidence of the connection of grit to outcomes such as performance and attrition is evaluated. Fifth, a short discussion of group differences in grit is presented. After a general summary of the empirical research in these first few sections and a discussion of what future research is needed, the final section will focus on whether and how grit might be a construct of value in relation to academic and workplace training preparedness.

What Is Grit?

Grit has been defined as “perseverance and passion for long-term goals” (Duckworth & Quinn, 2009, p. 166). Conceptually, grit is described as composed of two factors (consistency of interest and perseverance of effort). Gritty individuals are said to persevere following setbacks, adversity, or negative feedback. “Grit entails working strenuously toward challenges, maintaining effort and interest over years despite failure, adversity, and plateaus in progress. The gritty individual approaches achievement as a marathon: his or her advantage is stamina. Whereas disappointment or boredom signals to others that it is time to change trajectory and cut losses, the gritty individual stays the course.” (Duckworth, Peterson, Matthews & Kelly, 2007, p. 1088).

The basic dictionary definition of grit is “firmness of mind or spirit: unyielding courage in the face of hardship or danger” (Merriam Webster, 2014). Note this definition does not include a mention of long-term goals as does Duckworth’s. While some have suggested that people can have grit in reacting to the immediate context without thinking of the long term picture (e.g., Perlis, 2013), Duckworth’s definition emphasizes that grit is not about the short-term. As Eskreis-Winkler et al. (2014) state, “Grit is not just about working hard on tasks at hand but, rather, working diligently toward the same higher-order goals over extremely long stretches of time” (p. 21). Thus, the review in this paper is from the stance that the time frame of reference is an important conceptual element of grit.

Grit is discussed in Duckworth’s work as being important to sustained deliberate practice, which has been touted by Ericsson and Charness (1994) as key to expertise development (however, see Hambrick et al., 2014, for an opposing view on the role of practice in becoming an expert). Grit has also been described as enabling individuals to persist in repetitive or tedious behaviors needed to succeed (Silva, Eddington, Beaty, Nusbaum, & Kwapil, 2013). Arslan, Akin, and Citemel (2013) suggested that students with grit would continue working hard in classes even after failing tests and would look for new ways to understand material that is difficult. Duckworth and Quinn (2009) specifically noted this as a distinction between grit and Need for Achievement: Those high in grit do not care if a task is difficult – they stick to their goals even in the absence of positive feedback -- whereas those high in nAchievement would not persist if not achieving. Indeed, Duckworth et al (2007) stated that “individuals high in grit deliberately set for themselves extremely long-term objectives and do not swerve from them – even in the absence of positive feedback” (p. 1089). In line with a perspective that one can do well even if not achieving an end right away, Duckworth and Eskreis-Winkler (2013) argued that grit is related to growth mindset such that those higher in grit believe intelligence is malleable rather than fixed.

On the basis of this construct description, one would expect

- measures of grit to have two factors, interest and effort;
- examinations of grit to involve looking at behavior over long periods of time (e.g., years) and involve long term goals;
- grit to be associated with engaging in greater amounts of practice;
- experts in an area to be higher in grit than those that do not achieve expert status;
- individuals higher in grit to continue after negative feedback/events, whereas those lower in grit do not; and

- individuals high in grit to continue working on repetitive or boring tasks when those low in grit give up.

How Is Grit Measured?

The measurement of grit is prototypically via a short 8-item self-report scale, the Grit-S (see Appendix A for items). Earlier measure development work started from a pool of 27 items which led to the creation of a 12-item self-report scale, Grit-O. The Grit-O did not have great fit with the proposed two factor model (consistency of interest and perseverance of effort; CFI=.83, RMSEA = .11, Duckworth et al., 2007), which led to the development of the shorter form.

The Grit-S was developed by selecting items from 4 samples that had used the Grit-O and by looking at item-level correlations with outcomes (retention, GPA) and then eliminating two items from each of the two subscales that were most frequently below the median in predicting outcomes across the four samples. That is, items were eliminated that appeared the least predictive of outcomes. It is worth noting that 3 of the 4 eliminated items mentioned longer time perspectives (“My interests change from year to year,” “I become interested in new pursuits every few months,” and “I have achieved a goal that took years of work”) while only one of the remaining 8 items mentions longer time periods (i.e., “I have difficulty maintaining my focus on projects that take more than a few months to complete”). Thus, one question is whether the current measurement of grit, the Grit-S, focuses sufficiently on the long-term aspect as proposed conceptually.

Table 1 shows the Cronbach’s alpha reliabilities for the Grit-S and the Grit-O across samples. Test-retest reliability of the Grit-S was reported with a sample of high school students ($N=279$) in Study 4 of Duckworth and Quinn (2009) over a one-year period of $r_{xx} = .68$. Salles et al. (2014) found a test-retest correlation of .71 over 6 months with a sample of surgical residents ($N=141$). This limited evidence seems to support that the measure is reliable and relatively stable.

Factor analytic evidence to support the two underlying factors was not strong for the Grit-O but is relatively solid for the Grit-S. Table 2 shows supporting evidence for the factor structure. However, many studies do not report information on the subscales, and little work has been done to suggest their differential use. In one study that suggested how the subscales are distinct, Silva et al. (2013) showed that the subscales of grit had different relations with cardiac autonomic activity, with those high in perseverance showing higher sympathetic activity during the task and those high in consistency showing weaker sympathetic activity. They concluded that those higher in perseverance saw the task as more important. Overall, we know little about any distinctions in predictive value for the two factors.

Almost all research on grit uses either the Grit-O or Grit S measures, but there are a couple of exceptions. Study 3 of Duckworth and Quinn (2009) evaluated “other reports” of grit with a sample of 161 friend and family member reports on a modified Grit-S: correlations were .45 family member and self, .47 friend and self, and .37 family member and friend.

Duckworth’s lab website does have a “grit grid” which asks individuals to indicate their level of participation in different activities/hobbies/clubs and their level of achievement. While no further information on this self-administered measure was located, Robertson-Kraft and Duckworth (2014) used a 7-point rubric of a similar nature to assign grit scores to activities from teachers’ resumes. Coders scored two extracurricular activities from each resume (pilot work suggested no information was gained from scoring greater numbers of activities). One point was

assigned for participating in an activity for at least 2 years, an additional point for moderate achievement in the activity (holding a leadership position or receiving an award but not the highest possible), and 2 points given for highest achievement in the activity. Points for the two activities were summed for scores of 0-6. Appendix B illustrates the scoring scheme. Reported inter-rater agreement was $r = .82$ in Study 1 and $.85$ in Study 2.

One other way of assessing grit was identified. Vanderbilt University's Fisk-Vanderbilt Masters-to-Ph.D. Bridge Program uses an interview in admissions that purports to assess grit (Powell, 2013). Two faculty members score interview answers to questions about challenging experiences or obstacles encountered and how a candidate pulled through, using a grit scale. In a quote in the article on this measure, Duckworth expresses doubt about the usefulness of examining grit in high-stakes settings and that the interview might yield socially desirable answers. Oddly, she states that "The grittiest people are often the least willing to say they are gritty" (p. 473). This seems to run counter to the idea that a self-report measure of grit like the Grit-S is appropriate or useful.

In summary, grit is most typically measured by a short self-report measure. The overall scale has reasonable internal consistency and test-retest reliability, and seems to be composed of two factors. Subscale reliabilities appear to be more moderate and varied. Duckworth and colleagues (2007) acknowledge that the self-report grit measure is vulnerable to social desirability bias; however, no assessments of the extent of this bias were found.

What Evidence Supports the Construct Validity of Grit?

Considerable research has focused on how grit is conceptually distinct from related concepts. Of particular focus has been research relating grit to conscientiousness and to self-control (which itself is often viewed as a facet of conscientiousness). Studies have also reported on how grit relates to cognitive ability and other concepts. This research is summarized below.

Conscientiousness. There has been much discussion and investigation of whether grit is truly distinct from conscientiousness. Conceptually, Duckworth and colleagues originally argued for a clear distinction, noting that grit has to do with sustaining effort and interest in projects that take months or longer to complete, whereas individuals can be conscientious in the short term but not stick with things over a long time (Duckworth & Quinn, 2009). Although in a more recent paper coauthored by Duckworth (Eskreis-Winkler et al., 2014), grit is described as "one facet of Big Five Conscientiousness" and "a narrow facet of Conscientiousness" (p. 21), another recent coauthored paper (Robertson-Kraft & Duckworth, 2014) states "grit can be distinguished from conscientiousness" (p. 7). Others (e.g., Ivcevic & Brackett, 2014; Roberts, 2014) have referred to grit as a facet of conscientiousness, and the evidence summarized in the next few paragraphs seems to support that.

First, grit is highly correlated with Big Five measures of conscientiousness. Table 3 shows the correlations of the Grit-S with overall measures of conscientiousness (as well as with other Big Five personality variables) as reported across a number of studies. The correlation of grit and conscientiousness is typically quite high across samples (and would be higher once corrected for attenuation). The evidence that is typically presented to establish that grit is something more than conscientiousness is that it has been shown to add incremental prediction to outcomes when conscientiousness is controlled for. For example, in Study 2 of Duckworth and Quinn (2009) Grit-S predicted educational attainment (i.e., high school, some college, associate's...etc.) above age and the Big Five with a Nagelkerke R^2 of $.04$. In considering this evidence, though, it is important to note how conscientiousness is measured. Roberts,

Chernyshenko, Stark, and Goldberg (2005) have established that some conscientiousness measures emphasize industriousness and orderliness (e.g., NEO PI-R, AB5C) whereas others emphasize self-control (e.g., MPQ). In their study of 36 measures of conscientiousness, six facets emerged – impulse control, conventionality, reliability, industriousness, order and virtue – and they noted that no single existing Big Five measure of conscientiousness covers all of them. One question would be whether grit adds beyond measures of conscientiousness that contain “grit-like” facets such as industriousness and impulse control.

Second, empirical evidence supports the view that grit is a facet of conscientiousness. Roberts (2014) presents a table of correlations between the grit scales and six facets of conscientiousness (see Table 4) and concludes that grit is highly related to conscientiousness and its relation with other facets is similar to the intercorrelations found between those other facets. Indeed, Roberts, Lejuez, Krueger, Richards and Hill (2014) explicitly label grit as a subcomponent of conscientiousness.

There is a lack of clarity, however, over whether grit is another label for an established facet of conscientiousness or a distinct facet that has not been researched. For example, Ivcevic and Brackett (2014) state that “grit is a lower-level personality trait in the domain of Conscientiousness” (p. 29), noting that persistence has been identified as a facet of conscientiousness in a number of studies and that a number of the items in the grit scale are specifically about achievement goals. Duckworth and colleagues have noted that grit is similar to the facet of conscientiousness often called achievement striving, but claim it is distinct because of its emphasis on long-term goals and persistence in the face of setbacks. For example, in Duckworth et al. (2007) grit is described as differing from achievement orientation in that it emphasizes long-term stamina rather than short-term intensity.

Studies of the structure of conscientiousness typically put facets into two domains, orderliness and industriousness. There has been some support for viewing persistence as a more intersectional construct that bridges conscientiousness and ambition (a facet of extraversion) (DeRaad & Peabody, 2005; MacCann, Duckworth & Roberts, 2009), but persistence has generally been viewed as a form of industriousness (Roberts, et al., 2014). Thus, grit would fall into the industrious domain. However, the arguments of Duckworth and colleagues as well as Robert’s work (Table 4) suggest grit is distinguishable from tendency to work hard or aspiring to excellence, two facets of that subdomain of conscientiousness.

In summary, while Duckworth and colleagues have argued that grit is distinct from conscientiousness, their position appears to have softened. The empirical evidence does suggest that it would be reasonable to think of grit as a facet of conscientiousness, perhaps within the subdomain of industriousness.

Self-control. Duckworth and Gross (in press) note that although the terms self-control and grit are sometimes treated as synonymous, they are not. An empirical relation of .60 between the two has been reported (Duckworth et al., 2007; see also relation of .44 in Table 4). However, Duckworth and Gross argue that self-control and grit differ in the types of goals individuals “defend” and the time scale involved. They argue that self-control is about resolving a conflict between two action impulses associated with lower-level goals, whereas grit is about either not having competing superordinate goals or being able to actively suppressing rival superordinate goals. Accordingly, those high in grit, when faced with setbacks in a domain of interest, will develop new lower-order goals when others are blocked. Those with high grit maintain a focus on a high level goal over long stretches of time. It is important to note that

these are conceptual arguments, and studies evaluating the goals of gritty individuals were not located.

One study did look at behaviors that indicate a lack of self-control. Maddi et al. (2013) found grit to correlate $-.28$ with problematic internet usage, $.12$ with money conservation, $-.10$ with one measure and a non-significant relation with a different measure of compulsive buying. In general, however, these researchers concluded that hardiness was more of a protective factor with regard to problematic, addictive internet usage than grit. Another potential self-control behavior composite is that of rule violations (e.g., tardiness, incomplete homework) examined in a sample of high school students by Ivcevic and Brackett (2014). It was found to relate $-.18$ to grit.

In the section on outcomes more will be presented regarding behaviors of persisting and their association with grit, but overall the connections of grit to forming goals, sticking to goals and dealing with competing goals appear to be unaddressed. In general, both grit and self-control are considered facets of conscientiousness.

Cognitive Ability. Grit is assumed to not be strongly related to cognitive ability, although the empirical evidence actually does not provide a clear sense of the nature or magnitude of relations. Correlations have been reported of

- $-.20$ between Grit-O and SAT scores (Ivy League undergrads),
- $-.05$ and $-.08$ between Grit-O and SAT scores (West Point Cadets; Duckworth et al., 2007),
- $.08$ and $.06$ between grit coded from resumes and SAT (teachers, Studies 1 and 2, Robertson-Kraft & Duckworth, 2014),
- $.15$ between Grit-S and the Prairie State Achievement Test for Chicago Public School students in Eskreis-Winkler, et al., (2014),
- $.23$ between Grit-S and ACT scores reported by Strayhorn (2014),
- $.02$ of Grit-O with the Similarities subtest of the WISC-III in a sample of national spelling bee finalists (Duckworth et al., 2007), and
- $-.07$ Grit-S with ASVAB-GT in a ($N=677$) Army Special Forces sample (Eskreis-Winker et al., 2014, Study 1).

The variability in reported relations certainly reflects sampling error; the near zero relations, however, are almost all in samples where there was likely range restriction on both cognitive ability and grit, although information is not provided to enable corrections in most cases. Thus, further work with large samples with good variability on both grit and cognitive ability would be useful to verify that there is little relation between the two.

Other Constructs. Across studies, relations of grit with other constructs have been reported:

- gratitude $.25$ (Samson, Proyer, Coschi, Pedrini, & Ruch, 2011), $.44$ (Kleinman, Adams, Kashdan, & Riskind, 2013);
- life satisfaction $.34$ (Samson et al., 2011), $.17$ (Duckworth, Quinn, & Segliman, 2009), $.17$ (Hammer & Good, 2010);
- fear of being laughed at $-.22$ (Samson et al., 2011);
- subjective happiness $.27$ (Samson et al., 2011);
- meaning in life $.38$ (Kleinman et al., 2013);
- optimistic explanatory style $.32$ (Duckworth, Quinn, & Segliman, 2009);

- maximizing behavior v. satisficing .15 (Haran, Ritov, & Mellers, 2013);
- academic conscientiousness .49 (Eskreis-Winkler et al., 2014);
- school motivation .49 (Eskreis-Winkler et al., 2014);
- perceived teacher (.38), parent (.34) and peer (.42) support (Eskreis-Winkler et al., 2014);
- hours watching TV -.24 (at same time period) and -.22 (one year later) (Duckworth & Quinn, 2009; Study 4 of HS students);
- the subscale of perseverance of effort correlated .78 and consistency of interest correlated .79 with the metacognitive awareness inventory (Arslan et al., 2013);
- hardiness .60 (King & DeShon, 2014);
- procrastination -.68 (King & DeShon, 2014);
- resilience .46 (King & DeShon, 2014); and
- locomotion (self-regulation concerned with movement from state to state) .48 (King & DeShon, 2014).

What Outcomes Has Grit Been Linked To?

As grit is proposed to be related to persistence over the long term, most of the studies linking grit to outcomes focus on either retention or accomplishment. In this section, research on these two outcome categories is summarized, followed by a short discussion of research on persistence behaviors, an assumed meditational outcome between grit and retention or grit and accomplishment, and health outcomes which may connect to stamina and persistence.

Retention Outcomes

Academic Program Completion. Duckworth and Quinn's (2009) Study 5 of freshman West Point cadets looked at completion of the summer program – a point of early dropout -- and found Grit-S had incremental prediction beyond admittance score (a composite of high school rank, SAT, extracurricular activities and physical exercise) with change in Nagelkerke R^2 of .08. Maddie, Matthews, Kelly, Villarreal and White (2012) showed that grit was predictive of first year retention for West Point cadets, with those higher in grit twice as likely to complete the first year. Eskreis-Winkler et al. (2014, Study 1) found grit predicted completion of a 24-day "grueling" Army Special Operations Forces selection course with a Nagelkerke R^2 change of 1.84%, adding above ASVAB scores and physical fitness. In a study of 4813 Chicago high school students (Eskreis-Winkler et al., Study 3), grit, measured in the junior year, had a Nagelkerke change in R^2 of 0.50% in predicting graduation, while controlling for academic conscientiousness, standardized test scores, and other variables. Burkhart, Tholey, Guinto, Yeo, and Chojnacki (2014), in a study of 180 general surgery residents, found those lower in grit were twice as likely to report considering leaving, although no statistically significant relation with actual attrition was found (likely due to a small N attriting).

Turnover. In Study 2, Eskreis-Winkler et al. found that grit had a Nagelkerke change in R^2 of 1.27% in predicting staying in a sales job for six months, controlling for conscientiousness. In Study 1 of Robertson-Kraft and Duckworth (2014) a Cohen's d of .79 was reported in comparing the grit ratings from resumes of teachers who were retained or resigned at the end of a school year. Teachers with higher grit scores were more than twice as likely to be retained over the course of a year as those with lower scores.

Career Changes. If grit is associated with persistence over time, grit should be negatively associated with the number of lifetime career changes individuals make. In

Duckworth and Quinn (2009), the change in Nagelkerke R^2 for this outcome was only .01 beyond age and the Big Five. Eskreis-Winkler et al. (2014; Study 2) reported a .15 correlation between grit and years in sales.

Divorce. In Study 4 of Eskreis-Winkler et al. (2014), involving 6362 adults, “grit was association with a 17% increased odds of remaining married among men but was not associated with increased odds of remaining married among women” (p. 8).

In sum, research looking at grit as a predictor of retention or “sticking with it” in programs considered “difficult” to complete provides some support for the usefulness of grit, particularly for academic or military training programs. There are only a handful of studies that look at longer term outcomes such as job or career change.

Accomplishment Outcomes

Educational Attainment. Duckworth and colleagues have suggested that grit contributes to the completion of higher levels of education (Duckworth et al., 2007). Duckworth and Quinn (2009), Study 2, examined whether Grit-S predicted educational attainment (i.e., high school, some college, associate’s...etc.) above age and the Big Five; however, the change in Nagelkerke R^2 for grit in this study was only .01. Similarly, a .00 relation with years of schooling was reported in Study 1 of Eskreis-Winkler et al., (2014; $N = 677$ Army Special Operations Forces course enrollees), although there may have been little variance in educational levels in this sample. Duckworth et al. (2007) showed a connection to educational level (using Grit-O) with an eta squared of .05. Haran et al (2013) found the Grit –S measure to correlate with level of education (.19). It is important to note that examinations of educational level attained have not been longitudinal but cross-sectional. Overall, there is little evidence that those who go further in educational settings are necessarily grittier.

GPA. One assumption is that those with grit will persist in practice and studying in academic settings, and thus grit should have some relation with grades. Correlations of grit to GPA have been reported in several studies:

- .30 with same time GPA and .32 with GPA one year later in Duckworth & Quinn (2009) Study 4 of high school students (note that GPA in this study was an average across all subjects on 100 point scale);
- .25 with same time GPA in Study 3 of Duckworth et al (2007), with a relation of .34 once SAT scores were controlled;
- Maddie et al. (2012) showed grit to relate to first year performance scores for West Point Cadets. (effect size not reported);
- King and DeShon (2014) found a .13 correlation of grit with college GPA ($N = 467$);
- Strayhorn (2014) found Black male collegians’ grades to correlate with Grit-S .38, and their high school grades to correlate with grit .35;
- Ivcevic and Brackett (2014) found grit to correlate .18 with recognitions (ratings by faculty on work ethic and citizenship) but grit did not explain additional variance beyond conscientiousness in predicting GPA, achieving honors, or in recognitions ($N = 213$ private high school students); and
- college GPA and grit rated from resumes correlated .16 in Robertson-Kraft and Duckworth (2014) Study 1 and -.09 in Study 2.

In summary, grit does appear to relate to academic performance. Once again, these are cross-sectional studies and do not speak to any causal relation.

Teacher Effectiveness. Grit related to teaching effectiveness of novice teachers, with teachers one standard deviation higher in grit 31% more likely to outperform the less gritty (Duckworth, Quinn, & Seligman, 2009), with effectiveness based on students' normative academic performance one year later. In Robertson-Kraft and Duckworth (2014), similar findings for relations of novice teacher grit with effectiveness one year later were found. Specifically, a Cohen's d of 0.42 and 0.45 in grit was found between effective and ineffective teachers in two samples.

Advancement in Competition. Duckworth and Quinn (2009) Study 6 found that grit predicted the final round attained by spelling bee participants, with an incremental change in Nagelkerke R^2 of .02 beyond age and the Big Five (zero-order correlation with final round attainment of .16).

In sum, although grit does not appear to relate to educational attainment, it does appear to relate to performance in academic schoolwork and competitions as well as to the job performance of teachers.

Persistence Behavioral Outcomes

Hours practice. Duckworth and Quinn (2009), Study 6, found that grit related to accumulated lifetime self-reported spelling practice hours of .27. Duckworth, Kirby, Tsukayama, Berstein, and Ericsson (2011) found that grit was related to spelling performance (.17), deliberate practice (.31) and being quizzed on spelling (.19). That is, those high in grit engaged in more deliberate practice which then mediated the relation of grit to spelling performance (i.e., final round achieved in national bee).

Persistence at a Task. In Haran, et al. (2013), grit had a small (.02) incremental relation with persistence in acquiring information (number of times choosing to view objects before making an estimation of number of characters present on a screen). In an unpublished thesis study that looked at perseverance when repeatedly failing at a task (unsolvable number sequencing problems), Lowney (2013) did not find Grit-S to be related to performance level or task completion. The author does note that this is a short term task, and also might not be one for which participants had a goal.

In sum, given that grit is thought to have its influence on success through behaviors such as practicing more, repeating when failing, etc..., there are surprisingly few studies evaluating these relations. As noted, short term tasks such as in lab studies are not really appropriate. Individual goal commitment seems like it should be a boundary condition, although it is not assessed in these studies (i.e., even in the case of National Spelling Bee finalists, should we assume that participants in the study have a personal passion for spelling?) Overall, further work assessing persistence behavior is needed.

Health and well-being outcomes

Burnout. Salles, Cohen, and Mueller (2014) reported relations of grit and burnout (higher grit associated with lower burnout), with the relation with the emotional exhaustion and depersonalization subscales but not the personal accomplishment subscale.

Suicidal Ideations. Suicidal ideations might be viewed as an indicator of not being willing to persist when faced with adversity. Kleinman et al. (2013) found grit to correlate $-.28$ with suicidal ideations and $-.35$ with depressive symptoms, all measured 4 weeks later. However, grit was only negatively predictive of suicidal ideations when considered in combination with high levels of gratitude.

Health and Exercise Behaviors. Exercise, diet and other forms of health maintenance are often described as behaviors requiring persistence in order to achieve results. Reed, Pritschet and Cutton (2013) found grit to relate to moderate and high intensity stages of exercising but not to low intensity (i.e., initiating an exercise program), even after controlling for conscientiousness. Roberts (2014) reported a correlation of $.26$ between grit and wellness maintenance behaviors, $.35$ with accident control behaviors, $-.29$ with traffic risk behaviors, $-.07$ with substance risk behaviors and $.25$ with overall health behaviors ($N = 1,442$). Of note, the relations of grit with overall health behaviors and accident control both remained at $.15$ after controlling for six facets of conscientiousness.

In summary, grit does appear to relate to wellness behaviors as one would expect. Further investigation of how grit relates to sticking with exercise and diet regimes may enhance our understanding of the processes via which grit has its effects.

Are There Subgroup Differences in Grit?

Evaluation of a construct and its measurement should involve consideration of group differences. Differences in gender, ethnicity and age on measures of grit have been investigated.

Gender. Studies with Grit-S have generally found no gender differences in scores (Duckworth & Quinn, 2009, Study 2, Study 4, Study 6; Eskreis-Winkler et al., 2014, Study 2, Study 4; King & DeShon, 2014); studies with the resume rated grit have also found no relations (Robertson-Kraft & Duckworth, 2014). However, Samson et al. (2011) found grit-O to related $.14$ with gender (females scoring higher), and Eskreis-Winkler et al. (Study 3) found a similar $.14$ relation with the Grit-S. In terms of measurement equivalence across groups and settings, there is some evidence of equivalence across gender groups (Duckworth et al., 2009).

Ethnicity. Most studies located did not report information on how grit related to ethnicity. In the sales representative sample of Study 2 of Eskreis-Winkler et al. (2014), grit correlated $.10$ with being Black and $-.10$ with being White, while it was uncorrelated with being Asian or Hispanic. In their sample of 4812 public high school students (Study 3), the correlation was $.07$ with being Black and grit was unrelated to being White or Asian. Finally in their Study 4, only a $-.03$ relation with being White was reported and there was no relation with being Asian, Black, or Hispanic. King and DeShon (2014) reported a $-.07$ relation between grit and being white; and Robertson-Kraft and Duckworth (2014) reported $-.04$ and $-.06$ relations in their studies. No studies investigating measurement equivalence across subgroups were located.

Age. Studies have consistently found age relations such that older individuals report higher levels of grit:

- $.19$ in Duckworth and Quinn (2009), Study 2,
- $.16$ in Samson et al. (2011),
- $.31$ in Roberts (2014) ($N=1,442$), and

- .12, .20, and .12 in Eskreis-Winkler et al. (2014), studies 1, 2, and 4, respectively.

In sum, grit is not related to gender or ethnicity, but older individuals are grittier. Research on grit in cross-cultural contexts was not located (beyond one study in Turkey) so cultural equivalence of the measure and cultural differences in grit levels are unknown.

Critique of Empirical Evidence on Grit

In the initial discussion of the concept of grit at the beginning of this paper, several expectations based on the conceptual definition were noted. The empirical research that has been summarized supports some, but far from all, of those expectations. First, a reliable self-report measure of grit has been developed and does appear to have two factors as expected, interest and effort. Whether other report measures of grit have two factors was not investigated, and the biodata and interview measures of grit are not scored in terms of the two dimensions. Also, as noted earlier, most studies do not report subscale score relations with correlates, so it is difficult to develop a full understanding of how consistency of interest and persistence of effort relate to each other and to expected outcomes.

A second expectation is that examinations of grit should be over long periods of time (e.g., years) and involve long term goals. While many studies on grit are cross-sectional, there are a number that were longitudinal (e.g., Duckworth & Quinn, 2009, Study 4 had a 1-year lag, Study 5 had a 3-month lag, and Study 6 had a 1-month lag; Maddie et al., 2012 had a 1-year lag; Eskreis-Winkler et al., 2014, Study 1 had a 24-day lag, Study 2 had a 6-month lag, and Study 3 had a 1-year lag). However, it will take time to design and execute studies that look at very long term goals and that track individuals over multiple years; no study was located that extended beyond a one-year period.

A third implicit expectation is that grit should be associated with amount of practice, and as noted, this has been demonstrated with self-reported practice for spelling bee participants. A related expectation is that experts in an area will be higher in grit than those that do not achieve expert status. Beyond the spelling bee studies, work on grit has not looked at levels of practice or at the development of expertise, so more work to support these proposed theoretical relations is required.

Other expectations based on the conceptual definition of grit are that studies should examine persistence behaviors when tasks are repetitive or when individuals are faced with adversity or negative feedback. The only studies on persistence at a task are short term and cases where individuals are unlikely to have a long-term goal and passionate interest. Research that looks at retention and turnover is of course important in indicating the connection of grit to persistence, but ideally these investigations should also include assessments of individual's goal commitment and/or interest in the specific domain, as well as documenting adversity faced and negative feedback received. That is, if the theoretical underpinnings of grit are to be clearly supported, it should be clear that those high in grit end up with higher rates of retention despite facing adversity, rather than being able to complete a program or stay with a job without experiencing negative events or feedback.

Overall, the empirical basis for grit is growing, with a significant number of studies showing that grit does relate to retention and accomplishment outcomes as expected. Of note, in most cases the incremental validity from adding grit beyond overall conscientiousness is small (1–2% of variance); however, such effects can be practically meaningful. Research also

clearly indicates that grit is highly related to conscientiousness, and most researchers would label it as a facet of conscientiousness. The relation of grit to self-control probably warrants more investigation. It is interesting that Duckworth and colleagues (e.g., 2007) posit grit as an explanation for why “some individuals accomplish more than others of equal intelligence” (p1087) yet they do not control for intelligence in many of their incremental validity studies. Certainly more research with retention outcomes would be valuable, as would more research with non-self-report methods of assessing grit.

Potential Research Agenda

In addition to points noted in the critique of the empirical research, there are several other research directions that would aid our understanding of grit. First, given that grit is a facet of conscientiousness, general research on when broad v. lower-level traits are most predictive (Paunonen, Rothstein, & Jackson, 1999) should provide some guidance as to when grit will be more effective as a predictor than a broad measure of conscientiousness (Ivcevic & Brackett, 2014) and vice versa.

Second, the value of domain-specific measures of grit rather than the domain-general approach currently adopted should be explored (Duckworth & Quinn, 2009). Eskreis-Winkler et al. (2014) acknowledge that to establish whether grit is really something that is domain general, there is a need for research with the same individual across different settings. They provide the example that if grit is domain-general, it would predict who gives up in one setting (e.g., school) as well as in another (e.g., marriage).

Third, it seems that as grit is connected to goals, it should be a better predictor of success and retention in instances where individuals have choice than for success or retention outcomes in “required” settings (Ivcevic & Brackett, 2014). Ivcevic and Brackett note: “While students might be passionate about some subjects or activities, they are unlikely to be passionate about all subjects in high school. Thus, Grit might be a better predictor of achievement in self-selected narrower goals, such as performance in elective courses or extracurricular pursuits” (p. 33).

Fourth, Robertson-Kraft and Duckworth (2014) suggest that grittier individuals may use more active coping skills and seek more support when facing adversity, and perhaps are more proactive in seeking feedback and engaging in professional development; these relations should be investigated.

Perhaps the area most in need of research is whether interventions can be designed to either boost levels of grit or to assist the less gritty in persisting. There are no developed methods for training grit (Maddie et al, 2012). Indeed, Duckworth, Quinn and Segliman (2009) discussed that grit might not be susceptible to intervention, while at the same time suggesting that since grit is correlated with age perhaps it can be improved by intervention.

Besides the question of whether interventions related to grit make conceptual sense, there are a wide range of suggestions as to what the content of an intervention might be. VonCulin, Tsukayama and Duckworth (2014) suggest that interventions to increase grit can focus on particular motivational drives (desire for meaning in life, drive toward engagement) versus others (drive toward immediate pleasure). Arslan et al. (2013) suggest that educational programs aimed at improving metacognitive skills might improve grit. Strayhorn (2014) suggests that verbal persuasion can affect a “malleable trait” like grit, so parents and mentors should talk about the importance of perseverance. He also advocates structured opportunities for vicarious

experiences (e.g., watching others persevere). Duckworth and Gross (in press) suggest that as grit is not a skill that improves with practice but is more a motivational construct, interventions should target the antecedents of goal commitment such as thoughts about goal feasibility and desirability. Robertson-Kraft and Duckworth (2014) suggest interventions focus on learning to identify specific, changeable courses when facing adversity in a domain. Duckworth et al. (2007) suggest encouraging students to work not only with intensity but with stamina and let them know that excellence requires years and years of time on task. They also suggest rather than encouraging students to have breadth in coursework they might focus on depth. Overall, there are a lot of suggestions for how interventions might proceed but no investigation of these types of interventions in relation to grit levels.

Finally, Duckworth and Eskreis-Winkler (2013) mention that research is needed to determine if there are contexts where being gritty has some negative effects. Specifically, they speculate that grit might lead individuals to be vulnerable to the sunk-cost fallacy or to be less open to contradictory information, or to miss out on new opportunities because they are so focused on a goal. Thus, negative effects of grit (or curvilinear relations with outcomes) may exist.

How Grit Might Be of Value in Predicting Academic and Training Preparedness

Of interest to NAGB is the question of how useful grit might be in terms of understanding the readiness of 12th graders for academic and training settings beyond high school. Currently, the primary focus in assessing preparedness has been on levels of cognitive skills that indicate that students are ready for college or training programs without remediation. Grit is one exemplar of a motivational trait that might have particular importance in individuals being able to “stick with” college or a training program. In general, motivational constructs (e.g., self-regulatory skills) may help expand understanding of how and when some 12th graders who appear to have the skills (i.e., they “can do”) do not succeed (“will do”).

One difference between high school and post-secondary settings is that individuals have greater choice as to courses, programs of study, and careers. As noted earlier, grit should be of value when considering courses and training that individuals have chosen to engage in than for high school level work.

The Partnership for the 21st Century (P21; www.P21.org) skills framework offers a means of considering how grit fits into preparedness. The framework was developed to provide a common vision of the skills necessary for success in work and life. Relevant to this paper are Life and Career Skills, which include initiative and self-direction skills (manage goals and time, work independently, be self-directed learner) and also productivity and accountability skills (manage projects, produce results). The conceptual definition of grit, as well as the empirical research supporting it, suggests grit or similar facets of conscientiousness would be important predictors of these types of skills.

Grit can also be considered within the context of what have been called “achievement-relevant personality measures” (APMs; Briley, Domiteaux & Tucker-Drob, 2014). The research literature is replete with studies of intellectual investment, effort, and goal orientation; recently Briley et al. (2014) examined the commonalities underlying 36 APMs (e.g., goal orientation, procrastination, achievement striving, self-efficacy) and concluded that five factors – performance orientation, mastery orientation, self-doubt, effort and intellectual investment – explained the overlapping construct space of the measures. While Briley et al., (2014) did not

include grit as one of the APMs in their study, they note it could be evaluated as fitting into this nomological network.

While this paper focused specifically on grit, the broader question to ask is about the role of APMs in preparedness. That is, we have substantial empirical research that documents the importance of motivational constructs in academic success. Researchers have taken the step backwards to look at characteristics that predict academic motivation. In evaluating 12th grader preparedness, considering grit and other APMs may provide key information for students, educators, and policymakers on whether individuals are “motivationally prepared” to enter college or job-training programs.

References

- Arslan, S., Akin, A., & Citemel, N. (2013). The predictive role of grit on metacognition in Turkish university students. *Studia Psychologica*, 55, 311-320.
- Briley, D.A., Domiteaux, M., & Tucker-Drob, E.M. (2014). Achievement-relevant personality: relations with the Big Five and validation of an efficient instrument. *Learning and Individual Differences*, 32, 26-39.
- Burkhart, R.A., Tholey, R.M., Guinto, D., Yeo, C.J. & Chojnacki, K.A. (2014). Grit: a marker of residents at risk for attrition? *Surgery*, 1014-1022.
- Cobbett, W. (1829). *Advice to Young Men*. Retrieved from Google Books.
- Davis, V. (2014, January 9). *True grit: The best measure of success and how to teach it*. <http://www.edutopia.org/blog/true-grit-measure-teach-success-vicki-davis>.
- De Raad, B. and Peabody, D. (2005), Cross-culturally recurrent personality factors: analyses of three factors. *Eur. J. Pers.*, 19: 451–474.
- Duckworth A.L., & Eskresi-Winkler, L. (2013). True grit. *APS Observer*, 26, April.
- Duckworth, A.L., & Gross, J. (in press). Self-control and grit: Related but separable determinants of success. *Current directions in psychological science*.
- Duckworth, A.L., Kirby, T.A., Tsukayama, E., Berstein, H., & Ericsson, K.A (2011). Deliberate practice spells success: Why grittier competitors triumph at the National Spelling Bee. *Social Psychological and Personality Science*, 2, 174-181.
- Duckworth, A.L., Peterson, C., Matthews, M.D., & Kelly, D.R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92, 1087-1101.
- Duckworth, A.L., & Quinn, P.D. (2009). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment*, 9, 166-174.
- Duckworth, A.L., Quinn, P.D., & Seligman, M.E.P. (2009). Positive predictors of teacher effectiveness. *The Journal of Positive Psychology*, 4, 540-547.
- Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, 49, 725-747.
- Eskreis-Winkler, L., Shulman, E.P., Beal, S.A., & Duckworth, A.L. (2014). The grit effect: Predicting retention in the military, the workplace, school, and marriage. *Frontiers in Psychology*, 5, 1-12.
- Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Deliberate practice: Is that all it takes to become an expert? *Intelligence*, 45, 34-45.

- Hammer, J.H., & Good, G.E. (2010). Positive psychology: An empirical examination of beneficial aspects of endorsement of masculine norms. *Psychology of Men & Masculinity, 11*, 303-318.
- Haran, U., Ritov, I., & Mellers, B. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgment and Decision Making, 8*, 18-201.
- Ivcevic, Z. & Brackett, M (2014). Predicting school success: comparing conscientiousness, grit, and emotion regulation ability. *Journal of Research in Personality, 52*, 29-36.
- King, D.D., & DeShon, R.P. (2014, May). *The quest for parsimony: further examination of the grit construct*. Presented at the Association for Psychological Science conference, San Francisco.
- Kleiman, E.M., Adams, L.M., Kashdan, T.B., & Riskind, J.H. (2013). Gratitude and grit indirectly reduce risk of suicidal ideations by enhancing meaning in life: Evidence for a mediated moderation model. *Journal of Research in Personality, 47*, 539-546.
- Lowney, P. (2013). *The effects of grit, conscientiousness and question order of unsolvable problems on subsequent performance*. Unpublished thesis. DBS eSource.
- MacCann, C., Duckworth, A., & Roberts, R. (2009). Empirical identification of the major facets of conscientiousness. *Learning and Individual Differences, 19*(4), 451-458.
- Maddi, S.R., Erwin, L.M., Carmody, C.L., Villarreal, B.J., White, M., & Gundersen, K.K. (2013). Relationship of hardiness, grit, and emotional intelligence to internet addiction, excessive consumer spending, and gambling. *The Journal of Positive Psychology, 8*, 128-134.
- Maddi, S.R., Matthews, M.D., Kelly, D.R., Villarreal, B., & White, M. (2012). The role of hardiness and grit in predicting performance and retention of USMA Cadets. *Military Psychology, 24*, 19-28.
- Merriam-Webster (2014, July 28). Retrieved from <http://www.merriam-webster.com/dictionary/grit>
- Paunonen, S. V., Rothstein, M. G., & Jackson, D. N. (1999). Narrow reasoning about the use of broad personality measures for personnel selection. *Journal of Organizational Behavior, 20*(3), 389-405.
- Perlis, M.M. (2013) 5 characteristics of grit—how many do you have? *Forbes.com*, 10/29/2013.
- Powell, K. (2013). On the lookout for true grit. *Nature, 504*, 471-473.
- Reed, J, Pritschet, B.L., & Cutton, D.M. (2013). Grit, conscientiousness and the transtheoretical model of change for exercise behavior. *Journal of Health Psychology, 18*, 612-619.
- Roberts, B. (2014). <http://bwrobertspubs.wordpress.com/conscientiousness/>.
- Roberts, B.W., Chernyshenko, O.S., Stark, S., & Goldberg, L.R. (2005). The structure of

- conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology*, 58(1), 103-139.
- Roberts, B.W., Lejuez, C., Krueger, R.F., Richards, J.M., & Hill, P.L. (2014). What is conscientiousness and how can it be assessed? *Developmental Psychology*, 50(5), 1315-1330.
- Robertson-Kraft, C., & Duckworth, A. (2014). True grit: Trait-level perseverance and passion for long-term goals predicts effectiveness and retention among novice teachers. *Teachers College Record* 116.3: 1.
- Salles, A., Cohen, G.L., & Mueller, C.M. (2014). The relationship between grit and resident well-being. *The American Journal of Surgery*, 207, 251-254.
- Samson, A.C., Proyer, R.T., Ceschi, G., Pedrini, P.P., & Ruch, W. (2011). The fear of being laughed at in Switzerland: Regional differences and the role of positive psychology. *Swiss Journal of Psychology*, 70, 53-62.
- Schmitt, N. (2012). Development of rationale and measures of noncognitive college student potential. *Educational Psychologist*, 47(1), 18-29.
- Silva, P.J., Eddington, K.M., Beaty, R., Nusbaum, E.C., & Kwapil, T.R. 2013. Gritty people try harder: Grit and effort-related cardiac autonomic activity during an active coping challenge. *International Journal of Psychophysiology*, 88, 200-205.
- Strayhorn, T.L. (2014). What role does grit play in the academic success of Black male collegians at predominantly white institutions? *Journal of African American Studies*, 18, 1-10.
- Tough, P. (2013). *How children succeed*. New York, NY: Random House.
- VonCulin, K.R., Tsukayama, E., & Duckworth, A.L. (2014) Unpacking grit: Motivation correlates of perseverance and passion for long-term goals. *The Journal of Positive Psychology*, doi 10.1080/17439760.2014.898320

Table 1

International Consistencies for Overall Grit, the Persistence of Effort Factor and the Consistency of Interest Factor

Reference	Sample	N	Grit	Effort	Interest
Duckworth et al. (2007) ^a	West point cadets ('08 class)	1218	.73	.60	.73
Duckworth et al. (2007) ^a	West Point cadets ('10 class)	1308	.76	.65	.74
Duckworth et al. (2007) ^a	National Spelling Bee	175	.80	.65.	.76
Duckworth et al. (2007) ^a	Ivy League undergrads	139	.83	.78	.79
Duckworth and Quinn (2009), Study 2	Adult web sample (06-07)	1554	.70	.77	.82
Duckworth and Quinn (2009), Study 4	High school students	279	.82 ; .84	--	--
Duckworth and Quinn (2009), Study 5	West point cadets ('09 class)	1248	.77	--	--
Duckworth and Quinn (2009), Study 6	National Spelling Bee finalists	190	.82	--	--
Duckworth et al. (2007), Study 1	Adult web sample (04-05)	1545	.85 ^b	.78	.84
Samson et al. (2011)	Adult sample	196	.61 ^b	--	--
Duckworth, Quinn, and Seligman (2009)	Teachers	390	.77	--	--
Maddi et al. (2013)	Undergraduates	425	.89 ^b	--	--
Kleiman et al. (2013)	Undergraduates	209	.80 ^b	--	--
Silvia et al. (2013)	Students/adults	40	.74	.52	.69
vonCulin et al. (2014), Study 1	Adults website 08-10	15874	.82	.70	.83
vonCulin et al. (2014), Study 2	Adult MTurk	317	.82	.68	.84
Arslan et al. 2013	Students	352	--- ^c	.79	.76
Duckworth et al. 2011	Spelling bee finalists	274	.82	--	--
Hammer and Good (2010)	Men, from listservs	250	.82 ^b	--	--
Reed et al. (2013)	University employees	1171	.79	--	--
Ivcevic et al. (2014)	HS students	213	.72 ^b	--	--
Strayhorn (2014)	Black male college students	140	.87	--	--
Eskreis-Winkler et al. (2014) Study 1	Army special ops course participants	677	.77	--	--
Eskreis-Winkler et al. (2014) Study 2	Sales representatives	442	.79	--	--
Eskreis-Winkler et al. (2014) Study 3	HS students	4813	.90 ^d	--	--
Eskreis-Winkler et al. (2014) Study 4	Adults	6362	.79	--	--

^a These samples were collected in Duckworth et al. (2007) but the reliabilities for the shortened scale are reported in Duckworth and Quinn (2009) as the original study used the Grit-O

^b Grit-O scale used

^c Turkish version; overall alpha not reported

^d Used only 4 items of grit-S

Table 2

Confirmatory Factor Analysis of Grit-S with Two First-Order Factors and a Second-Order Latent Factor called Grit

Reference	Sample	<i>N</i>	χ^2	CFI	RMSEA
Duckworth et al. (2007)	West point cadets	1218	106.36	.95	.061
Duckworth et al. (2007) ^a	West Point cadets	1308	135.51	.95	.068
Duckworth et al. (2007) ^a	National Spelling Bee	175	71.57	.86	.101
Duckworth et al. (2007) ^a	Ivy League undergrads	139	43.63	.93	.097
Duckworth and Quinn (2009), Study 2	Adult web sample	1554	188.52	.96	.098
Arslan et al. (2013)	Student	352	41.72	.93	.059
King and DeShon (2014)	Student	467	111.92	.79	.10

^a These samples were collected in Duckworth et al. 2007, but the analyses for the shortened scale are reported in Duckworth & Quinn (2009) as the original study used the Grit-O

Table 3
Correlations of Grit-S with the Big Five Personality Traits

Reference	Sample	Big 5 Measure	N	Conscientiousness	Neuroticism	Agreeableness	Extraversion	Openness	
Duckworth and Quinn (2009), study 2, overall scale	Adult web sample	BFI	1554	.77	-.40	.24	.20	.06	
				Interest	.64	-.32	.18	.12	-.02
				Effort	.74	-.42	.25	.26	.14
Duckworth and Quinn (2009), study 6	National Spelling Bee finalists	BFI	190	.70	-.28	.44	.12	.18	
Duckworth, et al. (2007), study 2 ^a	Adult web sample (2006)	BFI	706	.77	-.38	.24	.22	.14	
Ivcevic and Brackett (2014) ^a	HS students	BFI	213	.44	-.33	.19	n/a	n/a	
Eskreis-Winkler et al. (2014), Study 2	Sales reps	BFI	442	.64	.48 ^b	.39	.25	.19	
Eskreis-Winkler et al. (2014), Study 4	Adults	BFI	6362	.71	.33 ^b	.20	.21	.08	

BFI (John & Srivastava, 1999)

^a Grit-O used instead of Grit-S

^b Correlation reported as with emotional stability rather than neuroticism

Table 4

Correlations between Grit Scales and 6-Facet Model of Conscientiousness (reproduced directly from Roberts, 2014)

Facets	Interest	Perseverance	Grit
Order	.34	.37	.41
Virtue	.33	.37	.41
Traditionalism	.30	.28	.35
Self-control	.38	.36	.44
Responsibility	.47	.55	.59
Industriousness	.44	.62	.61
Overall Conscientiousness	.51	.58	.64

N = 1442

Attachment A

Items from the Grit-S

Consistency of Interest

1. I often set a goal but later choose to pursue a different one.
2. I have been obsessed with a certain idea or project for a short time but later lost interest
3. I have difficulty maintaining my focus on projects that take more than a few months to complete
4. New ideas and projects sometimes distract me from previous ones

Perseverance of effort

5. I finish whatever I begin
6. Setbacks don't discourage me
7. I am diligent
8. I am a hard worker

Attachment B

Rubric for Rating Grit from Resumes, with Example Profiles, Reproduced Directly from Robertson-Kraft and Duckworth (2014), Table 1

Summary of Rubric for Rating Grit From Biographical Data in Applicant Résumés

Grit Rating	Sample Profile	Explanation of Rating
0	No multi-year involvement in any college activities	
1	Member of the swim team for three years but did not advance or win an award; no other multi-year activities	1 pt for multi-year activity; No other multi-year activities
2	Member of the mock trial team for two years and won the most improved award; no other multi-year activities	1 pt for multi-year activity +1 pt for moderate achievement in that activity; No other multi-year activities
3	Member of a fraternity for three years but no leadership roles; Assistant manager at the local movie theatre for three years	1 pt for multi-year activity; 1 pt for multi-year activity + 1 pt for moderate achievement in that activity
4	Camp counselor at local summer camp for three years; Player on the volleyball team for three years and captain in her senior year	1 pt for multi-year activity; 1 pt for multi-year activity + 2 pts for high achievement in that activity
5	President of the student body for three years; treasurer for the Kite and Key Club for two years	1 pt for multi-year activity + 2 pts for high achievement in that activity; 1 pt for multi-year activity + 1 pt for moderate achievement in that activity
6	Member of the cross-country team for four years and voted MVP in senior year; Founder and President for two years of the University's Habitat for Humanity chapter	1 pt for multi-year + 2 pts for high achievement in that activity; 1 pt for multi-year + 2 pts for high achievement in that activity

APPENDIX F

GRIT: A USEFUL CONCEPT IN COLLEGE AND CAREER PREPAREDNESS? (PRESENTATION)

MICHIGAN STATE UNIVERSITY

Grit: A useful concept in college and career preparedness?

Ann Marie Ryan
Michigan State University

1

MICHIGAN STATE UNIVERSITY

Why this topic?



- Non-cognitive predictors in college admissions and employment contexts
- Purportedly related to both performance and retention
- Buzz

2

MICHIGAN STATE UNIVERSITY

What is Grit?





3

MICHIGAN STATE UNIVERSITY

Grit

"perseverance and passion for long-term goals"
(p166, Duckworth & Quinn, 2009)

"Grit entails working strenuously toward challenges, maintaining effort and interest over years despite failure, adversity, and plateaus in progress. The gritty individual approaches achievement as a marathon: his or her advantage is stamina. Whereas disappointment or boredom signals to others that it is time to change trajectory and cut losses, the gritty individual stays the course."
(p1088, Duckworth, Peterson, Matthews & Kelly, 2007)



4

MICHIGAN STATE UNIVERSITY

Why might NAGB be interested in grit?

- being "prepared" involves psychological or mental preparation as well as having the requisite skill set.
 - a consideration of characteristics that relate to "sticking with it" may have value.
 - for settings that can be intense (e.g., compressed time frame courses, short probationary periods for on-the-job training) or challenging (e.g., programs with high failure rates)



← →

MICHIGAN STATE UNIVERSITY

Outline of Presentation

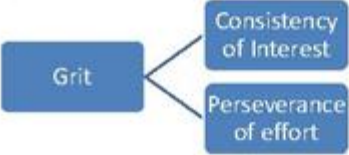
1. conceptual description
2. prototypical measurement
3. construct distinctiveness evidence
4. connection to outcomes: performance and attrition
5. group differences
6. relation to academic and workplace training preparedness.

← →

MICHIGAN STATE UNIVERSITY

Conceptual description

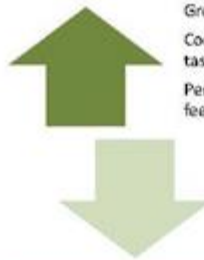
- - Over long periods of time
 - Involve long term goals



← →

MICHIGAN STATE UNIVERSITY

Conceptual description



- Greater practice
 - Continue with repetitive/boring tasks
 - Persevere following negative feedback or setbacks
- Less practice
 - Quit repetitive or boring tasks
 - Give up or lessen effort after negative feedback or when facing adversity

← →

MICHIGAN STATE UNIVERSITY

Grit-S

Consistency of Interest

- I often set a goal but later choose to pursue a different one.
- I have been obsessed with a certain idea or project for a short time but later lost interest
- I have difficulty maintaining my focus on projects that take more than a few months to complete
- New ideas and projects sometimes distract me from previous ones

Perseverance of effort

- I finish whatever I begin
- Setbacks don't discourage me
- I am diligent
- I am a hard worker

17

MICHIGAN STATE UNIVERSITY

Other measurement methods

- Grit-O (long form – 12 items)
- Grit grid
- Resume rubric
- Interview

18

MICHIGAN STATE UNIVERSITY

Grit-S Psychometrics


- Overall internal consistency .70 or higher
 - Subscale internal consistencies more in .60-.70 range
 - 2 factors supported but their intercorrelation seldom reported (.59)
- Test-retest reliability suggests relative stability
- Vulnerability to socially desirable responding?

19

MICHIGAN STATE UNIVERSITY

Grit v. Conscientiousness

- Grit has to do with sustaining effort and interest in projects that take months or longer to complete; individuals can be conscientious in the short term but not stick with things over a long time
 - ⇒ (Duckworth & Quinn, 2009).
- Incremental validity evidence
- Facet of C?
 - Industriousness




20

MICHIGAN STATE UNIVERSITY

Grit and Persistence behaviors

- Hours practice
- Persistence at a task



←

17

MICHIGAN STATE UNIVERSITY

Grit and health and well-being

- Burnout (-)
- Suicide ideation (-)
- Health and exercise behaviors



←

18

MICHIGAN STATE UNIVERSITY

Subgroup differences





←

19

MICHIGAN STATE UNIVERSITY

Critique of Evidence

- Two factors
 - Intercorrelation? Value in prediction?
- Long term goals
 - Some longitudinal studies, but not very long term
- Amount of practice and development of expertise
 - Beyond spellers
- persistence behaviors when tasks are repetitive
- persistence when individuals are faced with adversity or negative feedback.

←

20

MICHIGAN STATE UNIVERSITY

Other Research Needs

- Broad v. facet use in prediction
- Domain specific v. domain general
- Self-selected v. assigned goals
- Feedback seeking and support seeking
- Interventions
- Negative effects of grit

← 21 →

MICHIGAN STATE UNIVERSITY

Why of interest to NAGB

- Persistence as an outcome of importance
 - Prepared to persist
- Post-secondary settings are chosen
 - Self-selected goals
- Fit with P21 framework
 - initiative and self-direction skills
 - manage goals and time, work independently, be self-directed learner
 - productivity and accountability skills
 - manage projects, produce results

← 22 →

MICHIGAN STATE UNIVERSITY

Why of interest to NAGB

- Achievement relevant personality measures (APMs) and preparedness
 - Briley, Domiteaux & Tucker-Drob, 2014

Performance Orientation

Mastery Orientation

Self-doubt

Effort

Intellectual Investment

← 23 →

APPENDIX G

RELATING NAEP TO COMMERCIAL OFF-THE-SHELF MEASURES (PAPER)

**Nancy T. Tippins
Valtera Corporation
August, 2014**

Introduction

One of the challenges of NAGB is to link the NAEP test scores to preparedness for job training. Past research on NAEP suggests that some test items are not related to success in job training and concludes that the NAEP framework for math and reading is broader than that required for many job training courses. At the same time, there are likely other knowledge, skills, abilities, and other characteristics (KSAOs) that are related to preparedness in specific fields and jobs. In particular, “soft skills” such as perseverance, determination, conscientiousness, motivation for continuous learning, service orientation, safety awareness, and teamwork have not been factored into assessments of preparation for training.

Many employers hire large numbers of entry level workers and train them extensively in programs that have been developed specifically for the organization and its unique set of products, services, data systems, etc. For many employers, there are no educational programs that meet their needs outside of their own proprietary training programs. Consequently, these employers need to hire individuals who possess foundational skills that allow them to acquire the knowledge and skills necessary to perform the job. Even when occupationally specific training programs are available and taught through education and training institutions, including community colleges and vocational-technical schools, students are expected to have basic skills for learning by the time they enroll in the training courses.

This research proposal incorporates multiple measures of both the job and the individuals who apply, who are hired, and who are trained, and focuses on relating NAEP scores to scores on commercially available tests that are used for employee selection, as well as scores on an organization’s training tests that indicate who has successfully completed the training or certification assessments that identify those who perform at certain levels. In addition to evaluating the relationship of NAEP scores to measures of entry-level pre-requisites and training success, this study would also identify job requirements through a systematic job analysis that defines the KSAOs necessary to perform the work, compare what NAEP measures and the level at which NAEP measures those constructs to job relevant KSAOs, and then highlight gaps between the two that indicate deficits in candidate preparedness.

Purpose

The purpose of this study is to explore the relationship between NAEP test scores and preparedness for job training courses in business and industry. This relationship will be evaluated based on several measures of job preparedness collected at three points in time: prior to employment, during training, and after training. The degree of job preparedness is operationalized as scores on measures of job relevant KSAOs prior to employment, performance in training, and certification, which can include evaluations based on job knowledge tests, work sample tests, and key performance indicators (KPIs). It is assumed that training begins immediately after the individual starts the job. Thus, there is no material change in basic skills between the time of hire and the time training begins.

Three questions will be addressed:

1. To what extent are NAEP scores related to success in job training and/or certification?
2. Are the levels of reading and math items in NAEP consistent with the levels necessary for success in job training and/or certification?
3. What are the KSAOs necessary for success in training that are not measured by NAEP?

Approach

Because of the effort required to validate and implement operational pre-employment tests, create effective training programs, and develop certification programs, this research should be based on an organization's existing program. The use of commercially available, off-the-shelf measures of the job and the individual rather than custom-developed tools will facilitate replication across organizations. There are a number of test publishers that sell a variety of tests that are used for employee selection. Many offer tests in different formats, including multiple choice tests, work samples, situational judgment inventories (SJIs) that may be written or in audio or animated formats. Many firms have batteries of tests that have been validated for a job in multiple studies across organizations, and these batteries are frequently accompanied by job analysis questionnaires (JAQs). While some of the JAQs are tailored for the specific job, many are generic, rely on "universal" competency models, and apply to broad classes of jobs or to virtually all jobs.

Obtaining cooperation for this kind of work from any business can be a challenge in light of pressures for profitability. Because the ideal site for the research would be a company that has already collected job analysis data, test data, training scores, and certification scores, the only data collection efforts that remain are the collection of NAEP scores and the linkage of the NAEP constructs to the KSAOs measured in the organizations selection, training, and certification programs. Although this effort would certainly require the company's contribution of time from employees who take NAEP and who coordinate the efforts to assemble the other study data, companies that are challenged to find capable employees for particular jobs are likely to have an interest in ensuring that future high school graduates will possess minimal KSAOs that enable them to complete training successfully.

In many respects, this proposal constitutes a highly ambitious study, and it is likely that any participating organization would request paring the number of measures collected. Although not necessarily germane to NAGB's purposes, it is also possible that a participating organization would desire to look longer range and include performance data and promotion information. Nevertheless, record keeping for many of the components is required if the employer wants to comply with federal EEO laws and thus are likely to be available. The major missing variable would be NAEP scores and the linkage of job-relevant KSAOs to NAEP constructs.

In this study, two approaches exist for assessing the relationship between NAEP and job preparedness. One approach to evaluating the extent of the relationship is to compare the constructs measured by NAEP with the KSAOs required for training success. This approach relies on a careful job analysis of entry level jobs to identify the KSAOs required by the jobs and a comparison of them to the constructs NAEP measures. An important outcome of this part of the study is the identification of KSAOs that are important to training and job success that are not measured by NAEP.

A particularly valuable participating organization would be one with multiple entry-level jobs with different selection requirements, training programs, and certification standards. Analysis of diverse jobs would highlight the common requirements across multiple jobs as well

as the job-specific requirements. Comparison of the common job requirements to the constructs assessed by NAEP could identify the areas that are important to many entry level jobs but are not measured.

The second approach involves statistical analysis of the relationships between scores on NAEP between scores on tests commonly used for pre-employment decisions as well as measures of training success and job proficiency. Tests used for employee selection often assess cognitive abilities and non-cognitive skills as well as work sample tests that require the integration of multiple KSAOs. In business and industry, training success is often assessed through multiple-choice job knowledge tests and evaluations of performance. Similarly, certification of skills is commonly achieved through both knowledge test, simulations or performance tests, supervisory ratings of performance, and objective KPIs.

Process

The process to execute this study has two components: 1) the job analysis and 2) the statistical analysis.

Job Analysis

In employment testing, a job analysis is used to identify the tasks performed by incumbents in the job and the KSAOs necessary to perform the job. In turn, the KSAOs are used to determine which constructs to measure prior to employment, shape the training program, and construct experimental performance measures. To comply with federal EEO guidelines regarding pre-employment selection, a systematic approach to identifying the job requirements is required (Uniform Guidelines on Employee Selection Procedures, 1978).

The methods for conducting job analyses are varied; however, many consist of multiple steps, including job observations, interviews with incumbents and/or their supervisors, focus groups, and job analysis questionnaires (JAQs). The purpose of these steps is to compose lists of tasks and KSAOs so that they can subsequently be rated in the JAQ by incumbents and/or supervisors to provide quantitative data regarding the criticality of both tasks and KSAOs.

Many organizations conduct their job analyses using a “fresh start” approach and begin the process with no information about the job. Other organizations rely on existing information about the job to facilitate the process. For example, an organization that is conducting a job analysis on a job that has been previously studied may begin with the task and KSAO lists from the previous job analysis, skip job observations and interviews, and use the focus groups to discuss what has changed about the job since the last study so that the task and KSAO lists in the JAQ are current.

An increasingly used approach in many organizations is to use JAQs provided by the test vendor. While best practices and concerns about legal defensibility indicate that the task and KSAO lists should be reviewed and amended as needed to ensure completeness and accuracy, many organizations are willing to accept the lists as they are, particularly when the job under study is a common one that has many incumbents, e.g., sales, first line supervisory, customer service, etc., and that has been frequently studied. The use of a standard JAQ has the advantage of facilitating comparisons across different jobs in the same or in different organizations.

Depending on the nature of the original job analysis, some additional work may be required. Because of past research that suggests that the reading and math components of NAEP are broader than the requirements of training programs for many entry-level jobs, clear, accurate definitions of the levels of math and reading job requirements will be necessary. JAQ rating scales are typically designed to collect information on the frequency and importance of tasks and the importance of the KSAOs and the degree to which each is needed at job entry. In most cases, the KSAOs required on Day One (i.e., the first day of the job prior to any training) are the same as those needed for successful completion of the training program. If this is not the case, then it may be useful to ask subject matter experts to directly rate the extent to which each KSAO is necessary for training.

As noted above, an important analysis to be conducted is the assessment of the overlap between the KSAOs that are measured in NAEP and those required for successful performance as well as the identification of important KSAOs that are not related to the constructs measured by NAEP. To do this, a sample of testing experts will be asked to rate the similarity of each NAEP construct to each important KSAO.

Because of the time, effort, and expense of developing and validating employment tests, training programs, and certification processes and because the job analysis underlies all three programs, the underlying assumption is that the job analysis has already been done.

Statistical Analysis

The next phase of the study involves evaluating the statistical relationships between NAEP scores and the various scores that might be collected from job applicants and later job incumbents, including pre-employment test scores, training scores, and certification test scores. Although not germane to the concept of job readiness, job performance ratings will be useful to organizations that need to validate their pre-employment tests, and such analyses may serve as an enticement to participation.

Employers use tests that measure different KSAOs in different ways. Often, organizations use both cognitive and non-cognitive measures to cover a larger subset of the entire domain of KSAOs required and to attempt to mitigate adverse impact. Some organizations attempt to minimize the time spent testing applicants in order to speed the staffing process and attract applicants by using very short tests that focus only on the most critical KSAOs that are necessary on Day One. Cognizant of their investment in training, other organizations focus on a broader array of KSAOs and use more tests and more testing time to minimize training failures.

Organizations collect data regarding training performance in many different ways. Some organizations record only capstone training evaluations; others test trainees throughout the training courses, keeping records of all the scores. All training data that are available should be considered. Researchers may find that NAEP scores are more strongly related to training data collected earlier in the process than those later in the process because early training scores are more dependent on basic skills and later training scores are dependent on the effort the student has made to master the material as well as his/her basic skills.

Certification within organizations may be based on a single test score, job performance score, or KPIs, or certification may be based on a combination of such measures. There may be different relationships of NAEP scores to such measures. For example, the relationship between NAEP scores and a certification test scores may be stronger than that between NAEP scores

and KPIs because KPIs are a result of broader set of factors that include motivation, supervision, job opportunities, etc. Again, it will be important to consider all the components and their relationship to NAEP.

In some situations, researchers will need to consider which training or certification test score to use when retests are allowed. To the extent that the need for a retest reflects more limited basic skills, the relationship of NAEP scores and the first test may be stronger because of the greater degree of variance in the test scores. In this study, NAEP scores are likely to be collected after the training or certification test scores. If basic skills continue to develop as a person goes through training, the true relationship between NAEP scores and training/certification criteria may be attenuated.

By necessity, the organization(s) participating in this study have already collected these data. The development and validation of a pre-employment testing program, training program, and certification program is a significant undertaking that is beyond the scope of this proposal.

Statistical analyses are primarily simple inter-correlations of the variables available. Subgroup analyses may lead to further understanding of the relationships. In addition, other analyses may be useful. For example, if there is evidence of substantial restriction of range in the research sample due to selection standards or training rigor, corrections should be considered.

Example

To clarify the concepts and planned research, an organization that uses off-the-shelf JAQs and tests for three job families, operations, customer service, and sales, and that has developed a training course and a certification program for each job family is offered as an example. A job analysis for the entry level job in each family that identifies the critical KSAOs at entry has been conducted. Tests that measure the critical KSAOs for each of the entry-level jobs have been identified and validated. The tests for each job include a measure of mental ability, a broad work-oriented personality test, and an animated SJI that is related to the specific job. An extensive training program has been developed for each job that includes training progress tests given each week and a final training mastery assessment that consists of two parts, a job knowledge test and a work sample tests. Scores on each are combined for a final training score. A certification program has also been created. To be eligible to sit for the certification exam, employees must have been on the job for six months, and met KPIs related to their job family. The certification process itself requires a passing score on a job knowledge test and acceptable ratings by their supervisors on all performance dimensions on five different randomly chosen customer interactions.

A common JAQ based on a universal competency model was used for all three jobs. A group of ten testing experts linked the KSAOs evaluated in the JAQ to the constructs measured by NAEP. In the process of making the linkages, the subject matter experts reviewed the level of KSAOs measured in NAEP and required by the job and determined if they were the same or different. Then, three lists were constructed:

1. KSAOs measured by NAEP and required in the job
2. KSAOs measured by NAEP but not required by the job
3. KSAOs required by the job but not required by NAEP
- 4.

KSAOs in the first group represent alignment between NAEP and job requirements. When the number of KSAOs required by the job and linked to NAEP constructs is large, NAEP should be

highly related to job preparation. The KSAOs in the second group represent KSAOs that NAEP measures that are irrelevant to the job, and the third group of KSAOs represents those KSAOs that are important to the job and not measured by NAEP. To the extent that these KSAOs are required across three job families (as well as other families to be studied and other companies), these KSAOs represent areas in which NAEP might be expanded to better indicate job preparedness.

In the second part of this research effort, NAEP scores were correlated with the pre-employment test scores, all of the training test scores, and all of the certification assessment scores. Strong correlations between NAEP and the pre-employment test scores and training course scores indicate NAEP's relationship with job preparedness. Weaker correlations suggest a more tenuous relationship or the omission of a number of important, unmeasured variables in NAEP. If certification cannot occur until the employee has been on the job for six months or longer, the certification assessment represents a more distal criterion. Thus, one might suggest a weaker relationship because of the other variables that affect certification performance other than basic skills, i.e., motivation, training mastery, supervisory, opportunities' within the job, etc.

Challenges to the Study

Perhaps the greatest challenge to this study is finding an organization willing to participate. As noted above, the creation of pre-employment selection procedures, training programs, and certification programs is not a trivial undertaking. Few companies undertake an ambitious, integrated program, and those that have may not see value to the organization in deepening the understanding of NAEP and its relationship to the concept of job preparedness.

A related concern is the extent to which detailed records are kept within the organizations. Electronic applicant tracking systems have heightened the completeness and accuracy of pre-employment test scores, but training data in many organizations remains notoriously sketchy and incomplete. Even in companies with sophisticated record keeping, the last test score (not all test scores) are frequently kept.

Because NAEP scores are not used for pre-employment selection, low or non-existent correlations with selection test, training, or certification scores are not likely to harm the organization; however, if relationships among those three variables are explored, the organization may risk finding little support for the use of its selection procedures. Even if the pre-employment test has been successfully validated in the past, a new set of data based on a sample that is restricted by not only selection test scores but also training scores may yield different, less positive results. Statistical corrections based on the entire range of applicant data are possible, but are not commonly used on non-significant correlations.

The study as outlined here hinges on the availability of a job analysis based on a common JAQ that has been modified very little. Several potential problems related to this assumption could arise. The first problem that may arise is whether the organization has used the off-the-shelf JAQ with few modifications. The Uniform Guidelines specify steps for a defensible job analysis that allows modifications based on the job requirements and does not force compliance to a competency model developed elsewhere. A second potential issue is that the off-the-shelf JAQ does not specify reading and math levels in enough detail for a subject matter expert to determine whether or not the critical math or reading KSAO is a match to the constructs measured by NAEP. A third concern is whether the JAQs used across multiple jobs are similar enough to warrant conclusions about future inclusion of KSAOs that are not currently measured in NAEP. Obviously, the broadening of NAEP would be made on a number of studies

across multiple jobs and organizations; nevertheless, there should be some commonality across the studies to justify combining the results.

Some thought should go into the correlational analyses and their interpretations. Low correlations between NAEP scores and pre-employment test scores may not indicate that NAEP is not related to the job. Instead the weak correlation may indicate that the test score is dependent on a number of KSAOs, of which only a few are also measured in NAEP. The same finding and interpretation problems may also occur with training and certification scores. A training score may be partially dependent on basic skills but also contingent upon variables beyond the control of the trainee such as the skill of the trainer and opportunities to learn. Changes over time, particularly in training scores, merit some caution. Initial training scores may be more closely related to NAEP scores than later training scores because of the importance of other factors such as mastery of earlier materials, motivation, etc. that come into play as training progresses. A final concern in all applied studies such as this one is the reliability of the measures. Although NAEP is well-researched, training and certification tests often are not. Their reliability may be unknown and unacceptably low. Thus, weak correlations represent a poor measure rather than a weak relationship.

APPENDIX H

RELATING NAEP TO COMMERCIAL OFF-THE-SHELF MEASURES (PRESENTATION)

Relating NAEP to Commercial Off-the-Shelf Measures

Nancy T. Tippins

August 20, 2014

Purpose

- **Purpose**
 - Link NAEP scores to measures of job preparedness
- **Questions to be addressed**
 - To what extent are NAEP scores related to measures of job-preparedness?
 - To what extent are NAEP scores related to success in job training and/or certification?
 - Are the levels of reading and math items in NAEP consistent with the levels necessary for success in job training and/or certification?
 - What are the KSAOs necessary for success in training that are not measured by NAEP?

2

Approach

- **Identify organization that has**
 - Conducted a job analysis that identifies the KSAOs required to perform the job on Day One (i.e., prior to training)
 - Ensure that the level of math and reading is evaluated in the job analysis
 - Administers pre-employment tests and systematically retains the test data
 - Multiple-choice tests
 - Performance tests
 - Administers certification tests and systematically retains test data
 - Multiple-choice tests
 - Performance tests
 - Supervisor ratings of performance
 - Objective KPIs

3

Approach

- **Conduct appropriate analyses of the job analysis data**
 - Ask SMEs to rate similarity of NAEP constructs and KSAOs
 - Compare the constructs measured by NAEP to the KSAOs required to perform the job
 - Identify those KSAOs that are
 - Measured by NAEP and required in the job
 - Measured by NAEP but not required by the job
 - Required by the job but not required by NAEP
- **Statistically relate NAEP scores to pre-employment test scores, certification tests, and other associated measures**

4

Challenges to the Study

▪ **Willing organization(s)**

- Costs
- Confidentiality concerns
- Unintended consequences (e.g., adverse impact)
- Replicability across jobs and organizations

▪ **Sufficient measures**

- Inconsistencies in data storage
 - Retests
 - Validity and reliability of employment tests
 - Validity and reliability of certification tests, especially those that are "home-grown" and those that are based on supervisor ratings
 - Regional differences
 - Need to standardize KPI data
 - Restriction of range
 - Unmeasured variables
-

6