Competency-Based Testing and Credentialing: Addressing Inequity in Adult Education via State Policy Reform

Austin S. Jennings
eduCACE
United States

Abstract: Competency-based testing and credentialing (CBTC) initiatives aim to address inequity in adult education by fundamentally changing how states use GED®, HiSET®, and TASC™ test scores to award and withhold high school equivalency credentials. However, CBTC is inconsistent with how developers intend states to use those scores. Accordingly, it falls on states to evidence the validity of such interpretations. In the present study, I do just that – analyzing the extent to which GED®, HiSET®, and TASC™ test scores reflect interchangeable measures of academic attainment for the purpose of credentialing high school completion. Findings suggest high alignment between these tests across all sources and types of validity evidence. This finding – that CBTC-oriented use of GED®, HiSET®, and TASC™ test scores is valid – may appear simple, but the policy implications are complex. Considerations range from how states should handle non-authorized test scores to how they can support stakeholders' understanding of those scores.
Keywords: high school equivalency; standardized testing; credentialing; construct validity; equity in education; adult education

Pruebas y acreditación basadas en competencias: Abordando la inequidad en la educación de adultos mediante la reforma de políticas estatales
Resumen: Las iniciativas de pruebas y acreditación basadas en competencias (CBTC, por sus siglas en inglés) tienen como objetivo abordar la inequidad en la educación de adultos

al cambiar fundamentalmente cómo los estados utilizan las puntuaciones de los exámenes GED®, HiSET® y TASC™ para otorgar o retener credenciales de equivalencia de escuela secundaria. Sin embargo, CBTC es inconsistente con la intención de los desarrolladores sobre cómo los estados deben usar esas puntuaciones. En consecuencia, recae en los estados demostrar la validez de tales interpretaciones. En el presente estudio, analizo precisamente esto: hasta qué punto las puntuaciones de los exámenes GED®, HiSET® y TASC™ reflejan medidas intercambiables de logro académico para el propósito de otorgar credenciales de finalización de escuela secundaria. Los hallazgos sugieren una alta alineación entre estos exámenes en todas las fuentes y tipos de evidencia de validez. Este hallazgo –que el uso orientado a CBTC de las puntuaciones de GED®, HiSET® y TASC™ es válido– puede parecer simple, pero las implicaciones políticas son complejas. Las consideraciones van desde cómo los estados deberían manejar puntuaciones de pruebas no autorizadas hasta cómo pueden apoyar la comprensión de esas puntuaciones por parte de las partes interesadas.

Palabras clave: equivalencia de escuela secundaria; pruebas estandarizadas; acreditación; validez de constructo; equidad en la educación; educación de adultos

Testes e credenciamento baseados em competências: Abordando a inequidade na educação de adultos por meio da reforma de políticas estaduais

Resumo: As iniciativas de testes e credenciamento baseados em competências (CBTC, na sigla em inglês) buscam abordar a inequidade na educação de adultos ao mudar fundamentalmente como os estados utilizam os resultados dos testes GED®, HiSET® e TASC™ para conceder ou reter credenciais de equivalência ao ensino médio. No entanto, o CBTC é inconsistente com a intenção dos desenvolvedores sobre como os estados devem usar esses resultados. Assim, recai sobre os estados a responsabilidade de evidenciar a validade de tais interpretações. No presente estudo, faço exatamente isso – analiso até que ponto os resultados dos testes GED®, HiSET® e TASC™ refletem medidas intercambiáveis de desempenho acadêmico para fins de credenciamento de conclusão do ensino médio. Os resultados indicam uma alta consistência entre esses testes em todas as fontes e tipos de evidência de validade. Esse achado – de que o uso orientado ao CBTC dos resultados dos testes GED®, HiSET® e TASC™ é válido – pode parecer simples, mas as implicações políticas são complexas. As considerações variam desde como os estados devem lidar com resultados de testes não autorizados até como podem apoiar o entendimento desses resultados pelos stakeholders.

Palavras-chave: equivalência ao ensino médio; testes padronizados; credenciamento; validade de construto; equidade na educação; educação de adultos

Competency-Based Testing and Credentialing: Addressing Inequity in Adult Education via State Policy Reform

Across the United States, more than 27 million adults aged 18 and older have not earned a high school diploma or recognized equivalent (U.S. Census Bureau, 2024). For these individuals, high school equivalency (HSE) testing is a common pathway to high school completion, a milestone on their journey for continuing education, and a gateway to career opportunities.

In 2014, two developments fundamentally altered the HSE testing landscape. First, GED Testing Service (GEDTS, 2015) published the Fifth Series GED. In so doing, they transitioned to

computer-based testing, reconceptualized language arts as a single content area, aligned test items with contemporary K-12 content standards, eliminated compensatory standard setting recommendations, and recognized achievement beyond "high school equivalency" (GED® with Honors; GEDTS, 2015).

Second, and arguably more impactful, two test developers introduced alternatives to the GED Tests. Data Recognition Corporation (DRC, 2015) published the Test Assessing Secondary Completion (TASC) and Educational Testing Service (ETS, 2014) published the High School Equivalency Test (HiSET). In so doing, DRC and ETS reinforced many of the changes comprising the Fifth Series GED. They also provided states an opportunity to consider a wide range of equity-oriented social, political, and economic values at the core of HSE testing as a pathway to high school completion (Shaffer, 2015; Zinth, 2015).

States now had the opportunity to authorize multiple HSE tests based on factors ranging from academic rigor, standards alignment, and developer credibility to test format, price, and infrastructure (Shaffer, 2015; Zinth, 2015). States also began to grapple with early conceptualizations of *competency-based testing and credentialing* (CBTC) – understanding each HSE test as a way to measure a content area competency that may be combined with different measures of other competencies for the purpose of credentialing high school completion. States wondered, "If a candidate passed some Fourth Series GED tests in December 2013, should they have to re-demonstrate HSE based on newly-developed tests in January 2014?"

Test developers were split on this question. On one hand, GEDTS (2015) and DRC (2015) recommended that states only award credentials based on newly-developed tests. This perspective was grounded in belief that the Fifth Series GED and TASC set a more-rigorous standard for HSE than the Fourth Series GED (Shaffer, 2015; Zinth, 2015). On the other hand, ETS supported awarding credentials based on a combination of Fourth Series GED and HiSET test scores – a perspective that may have been more oriented toward business than test development. ETS' recommendation directly influenced states' initial test authorization and was promptly reversed the following year (Zinth, 2015).

Surprisingly, even though states considered credentialing high school completion based on a combination of Fourth Series GED and HiSET test scores, there is no evidence they considered doing so based on a combination of HiSET and either Fifth Series GED (hereafter *GED*) or TASC test scores. They did not openly wonder, "If a candidate passes some TASC tests before experiencing *interdeveloper mobility* – movement from one developer's HSE test battery to another's (Jennings, 2024) – should they be required to re-demonstrate HSE via GED or HiSET?" That is, until DRC (2022) announced they were discontinuing TASC. As if overnight, authorizing states began considering the equity-oriented implications of forcing some candidates to effectively restart their HSE journey if they could not pass the complete TASC test battery before it was gone.

This new consideration reflected a significant leap in states' conceptualization of CBTC and fell only a small step short of the bigger wondering, "If education stakeholders are expected to consider all HSE credentials equally, regardless of which test battery a candidate passes (Office of Postsecondary Education, 2014), should candidates be able to earn that credential by combining tests across those same batteries?" The present study is guided by, and designed to inform states' response to, this very question.

Research Perspective

Equity in education connotes structural and systemic conditions within institutions that produce and reproduce a fair and just relationship between opportunities, resources, and rules as

experienced by individuals those institutions serve (Adams, 1963; Arnautovic et al., 2022, Exec. Order No. 13985, 2021).

CBTC initiatives attend to one of many facets of equity in adult education: rules governing how states award and withhold HSE credentials based on HSE test scores. Traditionally, states legislate a unique set of rules for each test battery that determines who is eligible and what is required. As a result, candidates who start down one pathway before changing course need to effectively restart their HSE journey as though they had no pre-existing academic record. CBTC aims to address this by isolating each leg of that journey, allowing candidates to build a comprehensive academic record of their academic attainment across those same pathways. Under CBTC, a candidate who passed all TASC tests except TASC Mathematics before those tests were discontinued could earn their HSE credential by simply demonstrating mathematics competency through an alternate, state-authorized method.

While conceptually simple, advancing equity-oriented values through CBTC is complex. States need to review the rules and requirements governing each authorized pathway, reconceptualize those rules around underlying content area competencies, align those competencies across alternative pathways, and evaluate the extent to which pathway-specific requirements set comparable standards for each competency. That – specifically, as it pertains to the GED, HiSET, and TASC – is the objective of the present study.

When approaching research from an equity perspective, it is critical to position that research in relation to three considerations: whose experiences, in what space, and over what time (Stewart, 2013). With respect to whom, HSE testing as a pathway serves individuals whose identities disproportionately reflect a complex intersection of historically underserved communities in the United States. Racially, the odds of high school completion are 49.62% and 80.62% lower among Black and Hispanic individuals than non-Hispanic White, respectively (U.S. Census Bureau, 2023). Economically, the odds of living in poverty are 75.75% higher among those who have not completed high school than those who have but did not attend college (U.S. Census Bureau, 2022). Institutionally, the odds of high school completion are up to 78.65% lower among prisoners than noninstitutionalized citizens (Beatty & Snell, 2021; U.S. Census Bureau, 2017).

With respect to space and time, the present study focuses on post-2014 HSE testing in the United States. While states authorized the GED Tests for testing nonveteran adults as early as 1947 (Mullane, 2001), the present study focuses on three HSE tests formally recognized by the U.S. Department of Education: GED, HiSET, and TASC (Office of Postsecondary Education, 2014). Furthermore, although GED is administered in over 85 countries (GEDTS, 2019), the present study is not designed to evaluate claims regarding the extent to which GED is comparable to international alternatives for higher secondary certification (GEDTS, 2023).

Research Design

Test developers bear two responsibilities when publishing a new measure. First, they must define what test scores mean and how they are supposed to be used (American Educational Research Association [AERA] et al., 2014). Toward this end, DRC (2016, 2017), ETS (2018), and GEDTS (2018a) universally claim test scores measure candidates' content area-specific academic attainment for the purpose of credentialing high school completion, identifying strengths and weaknesses, and evidencing college and career readiness. Second, developers must evidence the validity of those claims through a mosaic of logical arguments, theoretical rationales, and empirical evidence (AERA et al., 2014; Kane, 2013; Messick, 1989; Moss et al., 2006).

When unintended meanings or uses of test scores arise, test users – not developers – bear the responsibility of evidencing the validity of those interpretations (AERA et al., 2014). CBTC reflects one of these cases.

On one hand, CBTC is consistent with what DRC (2016), ETS (2018), and GEDTS (2018a) intend test scores to mean. States intend to use test scores as a measure of candidates' content area-specific academic attainment. Accordingly, states do not need to reconsider the meaning of GED, HiSET, and TASC test scores. They have already done this.

On the other hand, CBTC is inconsistent with how DRC (2016), ETS (2018), and GEDTS (2018a) intend test scores to be used. Specifically, CBTC decouples content area tests from developer-specific batteries for the purpose of credentialing high school completion. That is, states intend to award HSE credentials based on complementary parts of multiple test batteries without requiring that candidates pass, or even attempt to pass, any developer-specific battery. Accordingly, states must evidence the extent to which GED, HiSET, and TASC tests may be used as interchangeable measures of content area-specific academic attainment for the fundamental purpose of credentialing high school completion.

Consistent with the *Standards for Educational and Psychological Testing* (hereafter *Standards*, AERA et al., 2014), the present study presents and evaluates the extent to which evidence supports the validity of the CBTC-oriented use of GED, HiSET, and TASC test scores.

Data Selection

The present study considers the corpus of validity evidence cited, developed, published, and maintained by or in coordination with DRC, ETS, and GEDTS. This includes current and archival technical manuals, content standards, test specifications, educational support materials, data analyses, norming and linking studies, as well as efficacy, implementation, research, and technical reports. As described in the *Standards* (AERA et al., 2014), "professional judgment is required to evaluate the extent to which existing validity evidence applies in the new situation … and to determine what new evidence may be needed" (p. 24). Toward this end, I requested additional data sources from test developers as deemed necessary. When test developers could provide that data – for example, auxiliary score scales – I explicitly cite and include it in the present evaluation. When test developers could not – for example, proprietary internal analyses – I explicitly note how its absence may affect the present evaluation.

Data Preparation

The present study comprises three stages of code development and application. Code development reflects both deductive and inductive approaches to qualitative analysis (Bingham & Witkowsky, 2022). Code application reflects a coder 1-led approach (Cole, 2024) subject to intercoder reliability via percent agreement due to the descriptive nature of the codebook, objective coding task, and independent code confirmability (Halpin, 2024).

First, I coded for sources of validity evidence. Contemporary scholars in educational measurement theorize validity as a unitary construct comprising multiple sources of evidence (Cronbach, 1971; Kane, 2013; Messick, 1989, 1995; Moss, 1992). Specifically, the *Standards* (AERA et al., 2014) – universally endorsed by DRC (2016), ETS (2018), and GEDTS (2018a) – advance five sources of validity evidence: test content, response processes, internal structure, relations to other variables, and consequences of testing. Accordingly, I attributed test developers' validity evidence to one of these five sources. These coded segments then served as the data set for subsequent stages of code development and application. Code confirmability was established via alignment with test

developers' independent validity evidence categorization. Coders demonstrated 100% agreement in code application.

Then, I coded for types of validity evidence. Contemporary scholars understand validity as not reliant on any specific form of evidence but rather a mosaic of convergent and divergent evidence inclusive of logical arguments, theoretical rationales, and empirical analyses (AERA et al., 2014; Messick, 1995). Accordingly, I iteratively re-coded each segment for emergent patterns in source-specific types of validity evidence as informed by the *Standards* (AERA et al., 2014). Code development concluded when all identified validity evidence was attributable to a mutually exclusive type of evidence (see Table 1). Code confirmability was established via alignment with types of evidence described in the *Standards*. Coders demonstrated 100% agreement in code application.

Table 1

Codebook for Sources and Types of Validity Evidence

Source	Type	Exemplar
Test content	Specification of content coverage	"Seventy-five percent of the texts in the [GED Reasoning Through Language Arts] exam are informational texts" (GEDTS, 2018a, p. 52).
	Alignment with content standards	"For the Geography domain, Social Studies content experts wrote the targets for the TASC Test based on the standards created by the National Council for the Social Studies and the National Council for Geographic Education" (DRC, 2020a, p. 8).
	Evaluation by content area experts	"Once the items have been reviewed internally, HiSET content specialists convene panels of educators to review the items and associated stimuli" (ETS, 2018, p. 15).
Response processes	Specification of cognitive response processes	"Evaluation: Judging the soundness or accuracy of scientific information or methods" (DRC, 2020a, p. 11).
	Analysis of cognitive response processes	"Response times per item type and content area were also calculated and assessed. This information was used to help determine the amount of testing time required for the operational test" (GEDTS, 2018a, p. 65).
Internal structure	Analysis of construct dimensionality	"Multi-factor CFAs with items loading on different content categories or subscores did not provide improved model fit for any of the subtests when compared to the results of the one-factor models." (ETS, 2018, p. 54).
	Analysis of differential item functioning	"Items that result in differential likelihoods of success for different subgroups are described as having differential item functioning (DIF). However, final judgment as to whether an item is biased toward one group over another is relegated to a panel of expert reviewers" (GEDTS, 2018a, p. 116).
Relations to other variables	Linking studies	"DRC developed this concordance resource to help program administrators, instructors, and students understand the relationship between TABE and the three High School Equivalency (HSE) tests: the TASC test, the GED®, and the HiSET®" (DRC, 2021, p. 1).

Source	Type	Exemplar
Consequences of testing	Norming studies	"Three stratifying variables were used to classify public school buildings across the nation: geographic region, district enrollment, and Title I status (and thereby socioeconomic status)" (ETS, 2014, p. 25).
	Establishing recommended cut scores	"In examining the performance of the 2013 Standardization and Norming Study participants at the 145 scaled score level, 72% would pass and 28% would be non-passers." (GEDTS, 2018a, p. 85).

Finally, I coded for the content area focus of validity evidence. CBTC frames test scores as interchangeable measures of content area-specific academic attainment for the purpose of credentialing high school completion. Accordingly, I re-coded each segment based on four a priori content areas: language arts, mathematics, science, and social studies. I then sub-coded language arts segments as either reading- or writing-focused to account for variation in DRC (2016), ETS (2018), and GEDTS' (2018a) perspective on the dimensionality of language arts test scores. Code confirmability was established via alignment with test developers' independent categorization of validity evidence. Coders demonstrated 100% agreement in code application.

Data Analysis

I addressed my guiding research question via descriptive and explanatory case study methodology (Priya, 2020; Yin, 2017). From a descriptive perspective, I present a parallel narrative of source-specific patterns in DRC, ETS, and GEDTS' validity evidence. From an explanatory perspective, I evaluate the extent to which that narrative supports those tests as comparable measures of content area-specific academic attainment. In this respect, comparability is operationalized as the degree of alignment between source-specific patterns in DRC, ETS, and GEDTS' validity evidence.

Of note, test-specific validity inferences are not a point of contention in the present study. I make no effort to evaluate the extent to which DRC, ETS, or GEDTS provide sufficient evidence to independently substantiate their intended meaning and uses of test scores. Instead, I focus on states' prospective, CBTC-oriented use of those scores as interchangeable measures of content area-specific academic attainment for the purpose of credentialing high school completion.

Findings

I structure findings in relation to the five sources of validity evidence advanced by the *Standards* (AERA et al., 2014) and universally endorsed by DRC (2016), ETS (2018), and GEDTS (2018a): test content, response processes, internal structure, relations to other variables, and consequences of testing. Each subsection begins with a description of the source, overview of patterns in related types of evidence, and general analysis of the extent to which that evidence substantiates GED, HiSET, and TASC test scores as comparable measures. Then, I provide a parallel narrative detailing that alignment in relation to each source-specific type of validity evidence.

Test Content

Test content-based validity evidence focuses on whether test items are relevant to, and representative of, the construct of interest (AERA et al., 2014; Messick, 1995). Emergent types of evidence include specification of content coverage, alignment with content standards, and evaluation by content area experts.

Evidence suggests DRC, ETS, and GEDTS cover comparable content after accounting for developer-specific reporting conventions. There are nominal differences in science and social studies test specifications, but those differences appear consistent with variation in traditional K-12 programs of study. All three developers subject test items to comparable expert evaluation, though only provide enough detail to support content area-neutral analysis of those efforts.

Specification of Content Coverage

Comparative analysis of content area test specifications addresses the question, "Do DRC, ETS, and GEDTS develop tests according to similar blueprints?" Findings suggest high alignment between developers' language arts and mathematics test specifications. However, GEDTS (2018a) does not specify the proportion of test items that cover writing-oriented content. Furthermore, variation in DRC (2020a), ETS (2021b), and GEDTS' (2018a) conceptualization of content categories limits comparative analysis of their relative coverage of those categories. Findings also suggest moderate alignment between developers' science and social studies test specifications. Table 2 presents a comparison of GED, HiSET, and TASC content coverage.

Table 2

GED, HiSET, and TASC Specification of Content Coverage

Content Area	GED	HiSET	TASC
Reading	Informational Texts (.75) Literary Texts (.25)	Informational Texts (.60) Literary Texts (.40)	Informational Texts (.75) Literary Texts (.25)
Writing	*Language conventions and usage*: Language conventions	*Selected response*: Organization of ideas (.22) Language facility (.43) Writing conventions (.35)	*Selected response*: Grammar and usage (.333) Capitalization, punctuation, and spelling (.208) Knowledge of language (.125) Text types and purposes (.167)
	Writing: Written analysis	*Constructed response*: Development of central position or claim Organization of ideas Language facility Writing conventions	*Constructed response*: Essay writing (.167)
Mathematics	Quantitative problem solving (.45) Algebraic problem solving (.55)	Numbers and operations on numbers (.19) Measurement and geometry (.18) Data analysis, probability, and statistics (.18) Algebraic concepts (.45)	Algebra (.26) Geometry (.23) Functions (.26) Number and quantity (.13) Statistics and probability (.12)
Science	Life science (.40) Physical science (.40) Earth and space science (.20)	Life science (.49) Physical science (.28) Earth science (.23)	Life sciences (.50) Physical sciences (.25) Earth and Space Sciences (.25)

Content Area	GED	HiSET	TASC
Social Studies	Civics and government (.50) United States history (.23) Economics (.13) Geography and the world (.13)	History (.35) Civics and government (.35) Economics (.20) Geography (.10)	United States history (.25) World history (.25) Civics and government (.20) Geography (.10) Economics (.20)

Note: Proportion of assessment tasks in parentheses, as available.

As detailed in relation to internal structure-based validity, the primary difference between language arts test specifications lies in developers' perspective on dimensionality. Specifically, DRC (2020a) and ETS (2021b) publish two language arts test specifications – one for reading and another for writing – while GEDTS (2018a) publishes a single blueprint that combines this content.

With respect to reading, all three developers emphasize informational texts (e.g., historical narratives, technical writing, nonfiction) in relation to literary texts (e.g., poetry, drama, fiction). This emphasis is slightly greater for DRC and GEDTS (p_{GED} = .75; p_{TASC} = .75) than ETS (p_{HiSET} = .60).

With respect to writing, DRC's (2020a) *conventions of standard English* and *knowledge of language* (p_{TASC} = .80) align with ETS' (2021b) *language facility* and *writing conventions* (p_{HiSET} = .78) in both content and coverage. These categories also align with GEDTS' *language conventions and usage* (p_{GED} not specified). Furthermore, DRC's *text types and purposes* (p_{TASC} = .20) aligns with ETS' *organization of ideas* (p_{HiSET} = .22) in both content and coverage. Again, these categories align with GEDTS' language assessment targets (p_{GED} not specified). Additionally, all three developers assess candidates' ability to develop, support, and communicate ideas while adhering to standard English writing conventions via constructed response items.

With respect to mathematics, GEDTS' *quantitative problem solving* (p_{GED} = .45) aligns with DRC's *numbers and quantities*, *geometry*, and *statistics and probability* (p_{TASC} = .48) as well as ETS' *numbers and operations on numbers*, *measurement and geometry*, and part of *data analysis, probability, and statistics* (range p_{HiSET}: .36-.55) in both content and coverage. Similarly, GEDTS' *algebraic problem solving* (p_{GED} = .55) aligns with DRC's *algebra* and *functions* (p_{TASC} = .52) as well as ETS' *algebraic concepts* and part of *data analysis, probability, and statistics* (range p_{HiSET}: .45-.64) in both content and coverage.

With respect to science, DRC (2020a), ETS (2021b), and GEDTS (2018a) cover similar content categories but vary in their relative emphasis on those categories. Consistent with DRC's (2016) evaluation of state graduation requirements, DRC (2020a) and ETS align in their emphasis on *life sciences* (p_{HiSET} = .49; p_{TASC} = .50) relative to *physical sciences* (p_{HiSET} = .28; p_{TASC} = .25). By contrast, GEDTS covers *life science* and *physical science* content equally (respectively, p_{GED} = .40). Despite this variation, all three developers align in their coverage of *earth and space sciences* (p_{GED} = .20; p_{HiSET} = .23; p_{TASC} = .25).

Like science, DRC (2020a), ETS (2021b), and GEDTS' (2018a) cover similar social studies content categories but vary in their relative coverage of history, civics, and government. GEDTS emphasizes *civics and government* (p_{GED} = .50) relative to history (p_{GED} = .23), with an explicit focus on *U.S. history*. ETS covers *civics and government* and history equally (respectively, p_{HiSET} = .35), with a joint focus on *U.S. and world history*. DRC delineates *U.S. history* from *world history*, covers them equally (respectively, p_{TASC} = .25), and emphasizes them relative to *civics and government* (p_{TASC} = .20). Despite this variation, developers are relatively aligned in their coverage of *economics* (p_{GED} = .13; p_{HiSET} = .20; p_{TASC} = .20) and *geography* (p_{GED} = .13; p_{HiSET} = .10; p_{TASC} = .10).

Alignment with Content Standards

Comparative analysis of content standards addresses the question, "Do DRC, ETS, and GEDTS develop test items according to similar content standards?" Findings suggest high alignment between language arts and mathematics test items but moderate-to-low alignment between science and social studies items. However, this lack of alignment more-closely reflects prominent critiques of the state of K-12 science and social studies standards (Helms et al., 2021; Stern et al., 2021) than a lack of comparability between HSE test items.

DRC (2015), ETS (2014), and GEDTS (2016a, 2016b) develop language arts and mathematics test items in relation to the Common Core State Standards (CCSS; Common Core State Standards Initiative [CCSSI], n.d.-a, n.d.-b). Of note, both DRC (2016) and ETS (2021b) transitioned toward a focus on College and Career Readiness Standards for Adult Education (CCRS; Pimentel, 2013) over time. That said, developers retain a high degree of alignment as CCRS are simply "a manageable set of the CCSS most indispensable for college and career readiness" (Pimentel, 2013, p. 5).

DRC (2015), ETS (2014), and GEDTS (2016c, 2020) also develop science and social studies items in relation to CCSS (CCSSI n.d.-a). However, CCSS are literacy-oriented standards and require that developers supplement them with other, content area-relevant standards (see Table 3).

Table 3

GED, HiSET, and TASC Alignment with Science and Social Studies Content Standards

Content Area	GED	HiSET	TASC
Science	A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas (National Research Council, 2012)	Self-published content category descriptors (ETS, 2021b)	Next Generation Science Standards (NGSS Lead States, 2013)
Social Studies	National Curriculum Standards for Social Studies: A Framework for Teaching, Learning, and Assessment (National Council for the Social Studies, 2010) National Standards for History: Basic Edition (National Center for History in the Schools, 1996)	Self-published content category descriptors (ETS, 2021b)	National Curriculum Standards for Social Studies: A Framework for Teaching, Learning, and Assessment (National Council for the Social Studies, 2010) National Standards for Civics and Government (Center for Civic Education, 1994) Voluntary National Content Standards in Economics (Council for Economic Education, 2010) Geography for Life: National Geography Standards (National Council for Geographic Education, 2012)

Note: Identified content area standards are in addition to CCSS (CCSSI n.d.-a) and CCRS (Pimentel, 2013), which advance literacy-oriented standards for science and social studies.

With respect to science, GEDTS (2020) develops test items in relation to National Research Council (2012) content standards whereas DRC (2020a) and ETS (2014) developed test items in relation to Next Generation Science Standards (NGSS Lead States, 2013). However, ETS (2018) subsequently transitioned to self-published content category descriptors derived from "local, state, and national guidelines...for high school curriculum" (p. 4).

With respect to social studies, DRC (2020a) and GEDTS (2016c) develop test items in relation to National Council for the Social Studies (2010) content standards. GEDTS also develops items in relation to National Center for History in the Schools (1996) content standards while DRC (2020a) aligns test items with Center for Civic Education (1994), Council for Economic Education (2010), and National Council for Geographic Education (2012) content standards. Consistent with their development of science test items, ETS (2018, 2021b) aligns social studies items with self-published content category descriptors.

Evaluation by Content Area Experts

Comparative analysis of expert evaluation addresses the question, "Do DRC, ETS, and GEDTS set similar standards for ensuring test items align with specified content standards?" Findings suggest high alignment between DRC (2016), ETS (2018), and GEDTS' (2018a) evaluative practices. However, unlike preceding types of evidence, all three developers describe predominantly content area-neutral evaluative practices.

DRC (2016), ETS (2018), and GEDTS (2018a) develop test specifications and items in conjunction with a range of education stakeholders including both internal and external evaluation. Specifically, GEDTS (2018a) draws on test item writers, specialists, and committees to ensure accuracy, content, coverage, quality, and fairness. ETS (2018) draws on specialists and panels of educators to iteratively review accuracy, content, relevance, and fairness. ETS also contracted with WestEd (2015) to conduct an independent analysis of language arts and mathematics standards alignment. DRC (2016) drew on internal and external specialists including the TASC Advisory Board to conduct sequential analyses of accuracy, content, context, difficulty, relevance, and fairness.

Response Processes

Response process-based validity evidence focuses on whether test items engage specified cognitive response processes (AERA et al., 2014; Messick, 1995). Emergent types of evidence include specification and analysis of cognitive response processes.

Evidence suggests DRC, ETS, and GEDTS specify comparable response processes and collect similar data related thereto. However, there is no evidence of whether or how developers used that data to evaluate the extent to which test items engage those response processes. That said, this reflects just one of many source-specific types of validity evidence and validity is not reliant on any specific such evidence (AERA et al., 2014).

Specification of Cognitive Response Processes

Comparative analysis of response process specification addresses the question, "Do DRC, ETS, and GEDTS intend for test items to engage similar response processes?" Findings suggest high alignment between developers' content area-specific enumeration of cognitive response processes (see Table 4). Of note, nominal within-content area differences are likely attributable to variation in developer-specific reporting conventions.

Across all four content areas, DRC (2020a, 2020b, 2020c, 2020d, 2020e, 2020f), ETS (2021b), and GEDTS (2016a, 2016b, 2016c, 2020) intend test items to engage candidates in analyzing, comprehending, evaluating, and reasoning.

Table 4

GED, HiSET, and TASC Specification of Cognitive Response Processes

Content Area	GED	HiSET	TASC
Reading	Analyzing Comprehending Evaluating Interpreting Reasoning Summarizing Synthesizing	Analyzing Comprehending Evaluating Interpreting Reasoning Summarizing Synthesizing	Analyzing Comprehending Evaluating Interpreting Reasoning Summarizing Synthesizing
Writing	*Language conventions and usage*: Analyzing Applying Comprehending Evaluating Reasoning	*Selected response*: Analyzing Applying Comprehending Evaluating Reasoning	*Selected response*: Analyzing Applying Comprehending Evaluating Reasoning
	Writing: Analyzing Communicating Reasoning Reflecting	*Constructed response*: Analyzing Communicating Reasoning	*Constructed response*: Analyzing Communicating Reasoning
Mathematics	Analyzing Applying Comprehending Evaluating Interpreting Reasoning Representing Summarizing	Analyzing Applying Comprehending Evaluating Interpreting Reasoning Representing Summarizing Synthesizing	Analyzing Applying Comprehending Evaluating Interpreting Reasoning Representing Summarizing
Science	Analyzing Applying Comprehending Evaluating Interpreting Reasoning Synthesizing	Analyzing Applying Comprehending Evaluating Interpreting Reasoning Synthesizing	Analyzing Applying Comprehending Evaluating Interpreting Reasoning
Social Studies	Analyzing Comprehending Evaluating Interpreting Reasoning Synthesizing	Analyzing Applying Comprehending Evaluating Interpreting Reasoning Synthesizing	Analyzing Applying Comprehending Evaluating Reasoning Summarizing

With respect to language arts, DRC (2020a, 2020c, 2020f), ETS (2021b), and GEDTS (2016b) also intend reading-oriented test items to engage candidates in interpreting, summarizing, and synthesizing. Furthermore, all three intend writing-oriented test items to engage candidates in applying, communicating, and reasoning. Additionally, GEDTS intends these items to engage candidates in reflecting.

With respect to mathematics, DRC (2020a, 2020b), ETS (2021b), and GEDTS (2016a) also intend test items to engage candidates in applying, interpreting, representing, and summarizing. Additionally, ETS intends these items to engage candidates in synthesizing.

With respect to science, DRC (2020a, 2020d), ETS (2021b), and GEDTS (2020) also intend test items to engage candidates in applying and interpreting. Additionally, ETS and GEDTS intend these items to engage candidates in synthesizing.

With respect to social studies, DRC (2020a, 2020e) also intends test items to engage candidates in summarizing. Furthermore, DRC and ETS (2021b) intend these items to engage candidates in applying. Additionally, ETS and GEDTS (2016c) intend these items to engage candidates in interpreting and synthesizing.

Analysis of Cognitive Response Processes

Comparative analysis of response process engagement addresses the question, "Do DRC, ETS, and GEDTS set similar standards for ensuring test items engage intended response processes?" Findings suggest alignment in cognitive response process engagement is indeterminate based on documentation currently available for peer review. However, available documentation references data that could serve as the basis for related empirical analysis (see DRC, 2016; ETS, 2018; GEDTS, 2018a).

Should states require evidence of response process engagement, extant literature advances varied methodological approaches including cognitive interviewing (Padilla & Leighton, 2017; Peterson et al., 2017), multilevel differential item functioning (Chen & Zumbo, 2017), and response time analysis (De Boeck & Jeon, 2019). However, it should be noted that all states have authorized at least one HSE test battery based on validity evidence that presently excludes any such analysis.

Internal Structure

Internal structure-based validity evidence focuses on whether test items are consistent with the construct of interest (AERA et al., 2014; Messick, 1995). Emergent types of evidence include analysis of construct dimensionality and differential item functioning.

Evidence suggests DRC, ETS, and GEDTS structure content area tests to measure comparable constructs of interest. Although developers do not agree on the dimensionality of language arts, evidence suggests a high degree of alignment in those constructs despite their difference in perspective. All three developers follow similar processes for analyzing and addressing test item fairness, though only provide enough detail to support content area-neutral analysis of those efforts.

Analysis of Construct Dimensionality

Comparative analysis of dimensionality addresses the question, "Do DRC, ETS, and GEDTS conceptualize test scores as measures of similar underlying constructs?" For example, GEDTS (2018a, 2020) aligns science test items with both CCSS (CCSSI, n.d.-a, n.d.-b) and National Research Council (2012) content standards. Accordingly, GED Science scores may measure candidates' academic attainment across three distinct dimensions: language arts, mathematics, and science. However, termed *unidimensionality*, all three developers typically intend for test scores to

measure a single construct of interest. Findings suggest high alignment between developers' conceptualization of dimensionality despite variation in their analytic approaches.

DRC (2016), ETS (2018), and GEDTS (2018a) present *item-total correlations*, a measure of the extent to which testers' individual item responses are associated with their total score (Guilford, 1953). All three developers flag test items indicative of poor discrimination for review and potential removal from their broader item pools.

ETS (2018) and GEDTS (2018a) also present *factor analyses*, a method of identifying relationships between test items within a set (Fabrigar & Kan, 2018; Randall & Jung, 2018). ETS (2018) analyzed unidimensionality via confirmatory factor analysis, finding within-content area multifactor models did not improve fit relative to single-factor models. Relatedly, GEDTS (2018a) analyzed unidimensionality via exploratory factor analysis, finding within-content area test items load on a single, dominant factor.

Of note, GEDTS (2018a) conceptualizes language arts as unidimensional whereas DRC (2016) and ETS (2018) conceptualize it as multidimensional, comprising distinct reading and writing constructs. However, both DRC and ETS report moderate intercorrelations between reading and writing test scores (respectively, $r = .63$; $r = .69$), suggesting notable overlap in those dimensions. Furthermore, ETS found model fit may be improved by combining some reading and writing factors but decided against those models as "estimated correlations among the reduced number of latent factors were still very high" (p. 54).

Analysis of Differential Item Functioning

Comparative analysis of differential item functioning (DIF) addresses the question, "Do DRC, ETS, and GEDTS set similar standards for analyzing and addressing differences in test item properties across subgroups of testers?" Findings suggest high alignment between developers' analysis of, and response to, DIF.

DRC (2016), ETS (2018), and GEDTS (2018a) analyze DIF in relation to gender, racial, and ethnic subgroups. All three developers flag test items indicative of poor functioning for further review by content area experts and fairness committees as well as potential removal from their broader item pools.

Relations to Other Variables

External variable-based validity evidence focuses on whether test scores are related to scores on other tests hypothesized to measure the same construct of interest (AERA et al., 2014; Messick, 1995). Emergent types of evidence include linking studies.

Evidence suggests DRC, ETS, and GEDTS follow comparable processes for evidencing external variable-based validity. However, detail regarding those studies is predominantly relegated to proprietary internal documentation. While each developer independently connects performance on their tests to a unique set of other measures, evidence directly supports comparative analysis of language arts and mathematics test scores.

Linking Studies

Comparative analysis of linking studies addresses the question, "Do DRC, ETS, and GEDTS set similar standards for ensuring test scores align with tests of similar constructs, difficulty, and reliability?" *Common population* concordances typically involve administering two tests to a single group of participants (Holland, 2007; Kolen, 2007). By contrast, *anchor study* concordances involve administering one unique and one common test to subgroups of participants. Findings suggest high alignment between developers' linking study methodology and findings. However, the proprietary

nature of these studies limits nuanced comparative analysis. That said, as detailed in relation to consequence-based validity evidence, alignment between developers is borne out in DRC's (2021) anchor test concordance study linking GED, HiSET, and TASC language arts and mathematics scores.

GEDTS (2018a) links their test scores with SAT and ACT scores via cut score transformation tables. ETS (2018; Klieger et al., 2022) links their test score with Iowa Tests of Educational Development (ITED) and ACT scores via minimum number of correct responses at the threshold for college and career readiness, content area-specific score correlations, and percent agreement in college and career readiness rates. ETS (2021a) also refers to data directly linking HiSET and GED test scores, but further detail is currently unavailable for peer review. DRC (2016) links their test scores with TABE 9 & 10 Level A and TerraNova Level 21 scores via content area-specific score correlations and transformation tables.

As noted above, DRC (2021) also linked GED, HiSET, and TASC language arts and mathematics scores via anchor test concordance using TABE 11 & 12 Level M, D, and A scores. As detailed in relation to consequence-based validity evidence, findings suggest all three developers recommend comparable HSE cut scores.

Consequences of Testing

Consequence-based validity evidence focuses on whether test score inferences support sound reasoning regarding the construct of interest (AERA et al., 2014; Messick, 1995). Emergent types of evidence include norming studies and establishing recommended cut scores.

Evidence suggests DRC, ETS, and GEDTS recommend comparable standards for measuring candidates' performance relative to that of traditional high school seniors. Despite apparent variation in developers' cut score recommendations, subsequent analysis of HSE candidate performance suggests developers recommend a similar standard for HSE in language arts and mathematics. However, the comparability of language arts-oriented evidence is limited by variation in developers' perspectives on its dimensionality.

Norming Studies

Comparative analysis of norming studies addresses the question, "Do DRC, ETS, and GEDTS set similar standards for relating testers' performance to that of high school seniors?" Findings suggest moderate-to-high alignment between developers' norming study methodology.

DRC (2016), ETS (2014), and GEDTS (2018a) norm test scores relative to high school seniors via *stratified sample*, a method of selecting participants based on population subgroups (DeYoreo, 2018). Specifically, developers weighted their final sample by strata including geographic region, community type, district-enrollment, metro status, and socioeconomic status to approximate the national population of high school seniors.

Two differences in developers' methodology bear mention to the extent participants' performance may be systematically related to differences in their effort and motivation. First, DRC (2016) and ETS (2014) selected participants via random sampling whereas GEDTS (2018a) employed a for-hire sampling model. Second, DRC and ETS collected norming data during participants' senior year of high school whereas GEDTS collected data after the school year concluded. To address these factors, developers accounted for rapid-guessing, item omission, testing time, and constructed response word count.

Establishing Recommended Cut Scores

Comparative analysis of recommended cut scores addresses the question, "Do DRC, ETS, and GEDTS' recommended HSE cut scores set a similar threshold for candidates' academic attainment?" Findings suggest moderate-to-high alignment between these thresholds. However, variation in developer-recommended language arts cut scores warrants a more-nuanced analysis.

DRC (2015) established recommended passing standards in coordination with CTB/McGraw-Hill. Across tests, recommended HSE cut scores reflect academic attainment greater than 30% of high school seniors.

ETS established recommended passing standards in coordination with panels of adult educators, content area experts, and a governing board of state-level decision makers (Tannenbaum & Reese, 2014). With the exception of HiSET Language Arts-Writing, recommended HSE cut scores reflect academic attainment equal to greater than 18 to 24% of high school seniors (ETS, 2014). For HiSET Language Arts-Writing, that threshold reflects academic attainment equal to or greater than 17.11 to 25.54% of high school seniors.

GEDTS (2015) established recommended passing standards via empirical and content-based analysis. Initially, recommended HSE cut scores reflected academic attainment equal to or greater than 28 to 31% of high school seniors, depending on content area (GEDTS, 2021). This aligned with DRC (2016) and the upper bound of ETS' (2014) recommendations. However, based on subsequent analysis, GEDTS (2018a) lowered their recommendation across all content areas. GEDTS' (2021) revised cut score recommendation reflects academic attainment equal to or greater than 12 to 18% of high school seniors, depending on content area. This predominantly aligns with the lower bound of ETS' recommendation, but is notably lower than that of DRC.

Although DRC's (2015) recommended cut score may appear to set a higher threshold for HSE than ETS (2014) or GEDTS (2021), DRC's (2021) anchor test concordance study suggests high alignment between those thresholds for language arts and mathematics. Specifically, DRC found the HSE cut scores for GED, HiSET, and TASC reading tests were comparable to a score of 536 on TABE 11 & 12 while the HSE cut score for all three mathematics tests were comparable to a score of 537.

Despite setting a comparable threshold for language arts achievement, nuanced variation in recommended cut scores reflects differences in related proficiency requirements. Whereas GEDTS (2018a) recommends a single language arts cut score, DRC (2016) and ETS (2018) recommend three cut scores: one for reading and two for writing. Furthermore, whereas DRC recommends three independent criteria, ETS recommends one dependent criterion. Due to this variation, candidates may pass GED Reasoning Through Language Arts without demonstrating written proficiency provided sufficient mastery of reading and language conventions (GEDTS, 2018b). Relatedly, candidates may pass HiSET Language Arts-Writing while demonstrating limited proficiency in selected response content provided sufficient mastery of constructed response content.

Discussion and Implications

CBTC addresses inequity by fundamentally changing how states use HSE test scores to award and withhold HSE credentials. Traditionally, states only award credentials to candidates who pass a complete test battery. Such policies treat candidates who experience interdeveloper mobility – whether voluntary or involuntary – as though they have no pre-existing academic record. CBTC aims to change this relationship by allowing candidates to combine HSE tests across authorized batteries and awarding credentials based on their complete academic record.

On the surface, CBTC initiatives appear rather common sense. If states award the same credential to candidates who pass different test batteries, why withhold that credential from candidates who pass complementary parts of those same batteries? However, this is not how DRC, ETS, and GEDTS intend states to use their test scores. Accordingly, it falls on states to evidence the validity of CBTC-oriented test score interpretations. The present study does just that.

With few exceptions, findings suggest high alignment between GED, HiSET, and TASC tests across all sources and types of validity evidence. All three developers provide overwhelmingly consistent evidence of their tests as valid measures of comparable constructs of interest. In other words, treating GED, HiSET, and TASC tests as interchangeable measures of content area-specific academic attainment is entirely consistent with considering any one of those tests to be a valid measure of HSE – which is precisely how developers intend states to use their test scores.

The primary finding of the present study is simple: CBTC-oriented uses of GED, HiSET, and TASC test scores are valid. The implications of that finding are complex. Toward this end, states should consider five questions when designing CBTC policies.

First, what academic competencies do you require? The present study focuses on four competencies traditionally measured by HSE test developers: language arts, mathematics, science, and social studies. Of note, findings suggest proficiency in language arts may be similarly evidenced by two tests, reading and writing, or combined into a single assessment. However, as discussed in relation to consequence-based validity evidence, states should consider the extent to which their language arts standard setting is consistent across those options. States should also consider academic competencies not traditionally measured by HSE test developers. For example, North Carolina's *Multiple Pathways to High School Equivalency* program includes a technology component (North Carolina Community Colleges, 2021). Toward this end, considering non-traditional competencies may help ensure candidates develop proficiencies critical to their continuing education and career opportunities.

Second, how do you handle non-authorized test scores? States consider a range of factors when authorizing HSE tests (Shaffer, 2015; Zinth, 2015), most of which have no bearing on the validity of test score interpretations. Traditionally, states handle out-of-state academic records via *interstate reciprocity*, honoring out-of-state test scores so long as they also authorize those same tests. If a candidate passes GED Mathematics in Pennsylvania before moving to New York, New York would honor that test score as they also authorize GED. If that same candidate passed HiSET Science before moving, New York would not honor that test score as they do not authorize HiSET. Findings do not suggest GED, HiSET, and TASC test scores are comparable, contingent on state authorization. They are simply comparable. Toward this end, findings suggest states should consider *interdeveloper reciprocity* – a policy whereby states honor GED, HiSET, and TASC test scores, regardless of independent authorization, so long as they were authorized by the candidate's state of residence at the time of testing.

Third, how do you handle other measures of academic attainment? The present study focuses on GED, HiSET, and TASC tests for a simple reason: test developers intend those scores to measure HSE. This has two benefits. First, I do not need to evidence the meaning of test scores. Test developers have already done that and states have already evaluated that evidence. Second, the evidence I need to evaluate CBTC-oriented uses of those scores is readily available. It is the same evidence developers provide to support their intended uses of test scores, just analyzed differently. When considering other adult education assessments – such as ACT WorkKeys, CASAS, or TABE in the case of Tennessee's *New Pathways* program (Tennessee Department of Labor and Workforce Development, 2024) – states bear a much greater burden and should exercise caution when extending the present methodology. States do not just need to evidence the comparability of those

test scores to GED, HiSET, and TASC. They need to evidence those scores as measures of HSE. Definitive claims supporting CBTC-oriented uses of those test scores will likely require extensive evidence that does not presently exist because, put simply, their developers never intended those tests to be used for credentialing high school completion.

Fourth, how do you handle compensatory score recommendations? GEDTS' (2015) elimination of a compensatory score recommendation is easily lost among all the other changes that came along with the Fifth Series GED. For the first time in over 70 years, GEDTS (2024) only recommended test-specific passing standards. They did not recommend a passing standard based on candidates' average or total score. DRC (2015) followed suit; ETS (2014) did not. Findings suggest ETS' recommended test-specific passing standards are comparable to those of DRC and GEDTS. However, those findings do not address how states can or should handle ETS' recommended battery-wide passing standard. Toward this end, education policy researchers should work with states to analyze options ranging from transforming test scores to eliminating such passing criteria.

Finally, how do you support stakeholders' understanding of multiple score scales? DRC (2016), ETS (2018), and GEDTS (2018a) maintain unique score scales, and for good reason. It facilitates stakeholders' understanding by limiting foreseeable conflation other measures (Kolen & Brennan, 2014; Petersen et al., 1989) including different HSE tests, measures of related constructs, and measures of general test performance. However, stakeholders are accustomed to only interpreting one scale at a time. When states require that candidates pass a complete HSE test battery, their transcripts necessarily include only test scores on that developer's scale. Under CBTC policies, stakeholders need to be able to interpret transcripts that include test scores reported on multiple different scales. Toward this end, education researchers should work with states to evaluate options for facilitating stakeholders' capacity to engage in data-driven decision making based on otherwise disparate score scales.

Limitations and Future Research

Two limitations of the present study bear mention. First, consistent with guidelines set forth in the *Standards* (AERA et al., 2014), the present study considers developers' existing body of validity evidence as well as some additional data sources they made available upon request. However, some data – such as proprietary internal analyses – was not accessible. Future research that includes such data may provide a more-nuanced perspective on the alignment of some source-specific types of validity evidence. Specifically, the extent to which developers set similar standards for evaluating response process engagement and ensuring test scores align with tests of similar constructs, difficulty, and reliability.

Second, the present study focuses exclusively on English language test formatting. Non-English language formatting was outside the scope of the present study and I did not request access to related validity evidence. That said, DRC, ETS, and GEDTS publish their tests in multiple language formats. Future research may extend the present methodology to comparatively analyze that evidence to inform how states handle cross-language HSE testing. Absent such research, a conservative approach would be for states to align CBTC policies with existing perspectives, allowing candidates to combine test scores across language formats when permitted through existing policies.

References

Adams, J. S. (1963). Towards an understanding of inequity. *Journal of Abnormal and Social Psychology*, *67*(5), 422–436. https://doi.org/10.1037/h0040968

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.

Arnautovic, I., Loew, D., Dogan, E., & Walton, E. (2022, August). *Equity Indicators Project: The Nation's "Equity Report Card."* 2022 NCES STATS-DC Data Conference, Virtual. https://nces.ed.gov/whatsnew/conferences/statsdc/2022/session_4.asp

Beatty, L. G., & Snell, T. L. (2021). *Survey of prison inmates: Profile of prison inmates, 2016* (No. NCJ 255037). Bureau of Justice Statistics.

Bingham, A. J., & Witkowsky, P. (2022). Deductive and inductive approaches to qualitative data analysis. In C. Vanover, P. Mihas, & J. Saldaña (Eds.), *Analyzing and interpreting qualitative data: After the interview* (pp. 133–146). SAGE Publications, Inc.

Center for Civic Education. (1994). *National standards for civics and government*. Author.

Chen, M. Y., & Zumbo, B. D. (2017). Ecological framework of item responding as validity evidence: An application of multilevel DIF modeling using PISA data. In B. D. Zumbo & A. M. Hubley (Eds.), *Understanding and investigating response processes in validation research* (pp. 53–68). Springer. https://doi.org/10.1007/978-3-319-56129-5_4

Cole, R. (2024). Inter-rater reliability methods in qualitative case study research. *Sociological Methods & Research*, *53*(4), 1944–1975. https://doi.org/10.1177/00491241231156971

Common Core State Standards Initiative. (n.d.-a). *Common Core State Standards for English language arts & literacy in history/social studies, science, and technical subjects*. Author. https://corestandards.org/wp-content/uploads/2023/09/ELA_Standards1.pdf

Common Core State Standards Initiative. (n.d.-b). *Common Core State Standards for mathematics*. Author. https://corestandards.org/wp-content/uploads/2023/09/Math_Standards1.pdf

Council for Economic Education. (2010). *Voluntary national content standards in economics* (2nd ed.). Author.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443–507). American Council on Education.

Data Recognition Corporation. (2015). *Technical manual: TASC test 2014*. Author.

Data Recognition Corporation. (2016). *TASC test technical report 2016*. Author.

Data Recognition Corporation. (2017). *A new TASC test performance level: Distinguished achievement benchmark*. Author.

Data Recognition Corporation. (2020a). *Objective structure: TASC test PQR*. Author.

Data Recognition Corporation. (2020b). *TASC Mathematics blueprint*. Author.

Data Recognition Corporation. (2020c). *TASC Reading blueprint*. Author.

Data Recognition Corporation. (2020d). *TASC Science blueprint*. Author.

Data Recognition Corporation. (2020e). *TASC Social Studies blueprint*. Author.

Data Recognition Corporation. (2020f). *TASC Writing blueprint*. Author.

Data Recognition Corporation. (2021). *TABE to high school equivalency (HSE) test concordance*. Author.

Data Recognition Corporation. (2022). *Data Recognition Corporation (DRC) - Notification*. TASC Test. https://tasctest.com/

De Boeck, P., & Jeon, M. (2019). An overview of models of response times and processes in cognitive tests. *Frontiers in Psychology*, *10*, 1–11. https://doi.org/10.3389/fpsyg.2019.00102

DeYoreo, M. (2018). Stratified random sampling. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (p. 1624). SAGE Publications, Inc. https://doi.org/10.4135/9781506326139.n671

Educational Testing Service. (2014). *HiSET® technical manual*. Author.

Educational Testing Service. (2018). *HiSET® technical manual*. Author.

Educational Testing Service. (2021a). *2020 Annual Statistical Report on the HiSET® Exam*. Author. https://hiset.ets.org/s/pdf/2020_annual_statistical_report.pdf

Educational Testing Service. (2021b). *HiSET® – test at a glance (TAAG) information brief*. Author.

Exec. Order No. 13985, 86 F.R. 7009 (2021).

Fabrigar, L. R., & Kan, M. P. H. (2018). Exploratory factor analysis. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 649–653). SAGE Publications, Inc. https://doi.org/10.4135/9781506326139.n252

GED Testing Service. (2015). *Technical manual: 2014 GED® test*. Author.

GED Testing Service. (2016a). *Assessment guide for educators: Mathematical reasoning*. Author. https://ged.com/wp-content/uploads/assessment_guide_for_educators_math.pdf

GED Testing Service. (2016b). *Assessment guide for educators: Reasoning through language arts (RLA)*. Author. https://ged.com/wp-content/uploads/assessment_guide_for_educators_rla.pdf

GED Testing Service. (2016c). *Assessment guide for educators: Social studies*. Author. https://ged.com/wp-content/uploads/assessment_guide_for_educators_social_studies.pdf

GED Testing Service. (2018a). *Technical manual: GED® test (Updated 2018 edition)*. Author.

GED Testing Service. (2018b, July). *What students need to know: GED® RLA's extended response*. https://ged.com/wp-content/uploads/Writing-What-We-Know-Using-ER-Scoring-to-Improve-Performance.pdf

GED Testing Service. (2019). *GED® international test program updates*. https://ged.com/insession/ged-international-program_march2019/

GED Testing Service. (2020). *Assessment guide for educators: Science*. Author. https://ged.com/wp-content/uploads/assessment_guide_for_educators_science.pdf

GED Testing Service. (2021). *GED scaled scores* [Dataset].

GED Testing Service. (2023). *Go to a university. Anywhere*. GED® Official Website. https://ged.com/en/

GED Testing Service. (2024). *Score scale and content descriptions for the GED® tests*. https://ged.com/score_scale/

Guilford, J. P. (1953). The correlation of an item with a composite of the remaining items in a test. *Educational and Psychological Measurement*, *13*(1), 87–93. https://doi.org/10.1177/001316445301300109

Halpin, S. N. (2024). Inter-coder agreement in qualitative coding: Considerations for its use. *American Journal of Qualitative Research*, *8*(3), 23–43. https://doi.org/10.29333/ajqr/14487

Helms, J., Nations, J., & Randall, D. (2021). *Climbing down: How the Next Generation Science Standards diminish scientific literacy*. National Association of Scholars.

Holland, P. W. (2007). A framework and history for score linking. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 5–30). Springer. https://doi.org/10.1007/978-0-387-49771-6_2

Jennings, A. S. (2024). *On the conceptualization of interdeveloper mobility* (Report No. RM-2024-01). eduCACE. https://doi.org/10.71180/cace015932

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. https://doi.org/10.1111/jedm.12000

Klieger, D. M., Williams, K. M., Bochenek, J. L., Ezzo, C., & Jackson, T. (2022). *Validating HiSET® tests as high school equivalency tests that improve educational, vocational, and quality-of-life outcomes* (No. RR–22-14). ETS. https://doi.org/10.1002/ets2.12359

Kolen, M. J. (2007). Data collection designs and linking procedures. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 31–55). Springer. https://doi.org/10.1007/978-0-387-49771-6_3

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). Springer. https://doi.org/10.1007/978-1-4939-0317-7

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Macmillan; American Council on Education.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *The American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Moss, P. A. (Fall 1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, *62*(3), 229–258. https://doi.org/10.3102/00346543062003229

Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Chapter 4: Validity in educational assessment. *Review of Research in Education*, *30*(1), 109–162. https://doi.org/10.3102/0091732X030001109

Mullane, L. (2001). *Bridges of opportunity: A history of the Center for Adult Learning and Educational Credentials*. American Council on Education.

National Center for History in the Schools. (1996). *National standards for history: Basic edition*. Author.

National Council for Geographic Education. (2012). *Geography for life: National geography standards* (2nd ed.). Author.

National Council for the Social Studies. (2010). *National curriculum standards for social studies: A framework for teaching, learning, and assessment*. Author.

National Research Council. (2012). *A framework for k-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press. https://doi.org/10.17226/13165

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press. https://doi.org/10.17226/18290

North Carolina Community Colleges. (2021). *Multiple pathways to high school equivalency (MPHSE) diploma program: Operating guidelines*. Author.

Office of Postsecondary Education. (2014). *Recognized equivalent of a high school diploma* (No. GEN-14-06). United States Department of Education. https://fsapartners.ed.gov/sites/default/files/attachments/dpcletters/GEN1406.pdf

Padilla, J.-L., & Leighton, J. P. (2017). Cognitive interviewing and think aloud methods. In B. D. Zumbo & A. M. Hubley (Eds.), *Understanding and investigating response processes in validation research* (pp. 211–228). Springer. https://doi.org/10.1007/978-3-319-56129-5_12

Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 221–262). Macmillan; American Council on Education.

Peterson, C. H., Peterson, N. A., & Powell, K. G. (2017). Cognitive interviewing for item development: Validity evidence based on content and response processes. *Measurement and Evaluation in Counseling and Development*, *50*(4), 217–223. https://doi.org/10.1080/07481756.2017.1339564

Pimentel, S. (2013). *College and Career Readiness Standards for Adult Education*. Office of Vocational and Adult Education, U.S. Department of Education.

Priya, A. (2020). Case study methodology of qualitative research: Key attributes and navigating the conundrums in Its application. *Sociological Bulletin*, *70*(1), 94–110. https://doi.org/10.1177/0038022920970318

Randall, J., & Jung, H. J. (2018). Confirmatory factor analysis. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 365–370). SAGE Publications, Inc. https://doi.org/10.4135/9781506326139.n140

Shaffer, B. (2015). *The changing landscape of high school equivalency in the U.S.: Options, issues, and improvement strategies*. Center for Postsecondary and Economic Success. https://www.clasp.org/wp-content/uploads/2022/01/The-Changing-Landscape-of-High-School-Equivalency-in-the-U.S.-Final.pdf

Stern, J. A., Brody, A. E., Gregory, J. A., Griffith, S., Pulvers, J., Griffith, D., & Northern, A. M. (2021). *The state of State standards for civics and U.S. history in 2021*. Thomas B. Fordham Institute.

Stewart, F. (2013). *Approaches towards inequality and inequity: Concepts, measures and policies* (Innocenti Discussion Papers). UNICEF Office of Research - Innocenti. https://doi.org/10.18356/56df5bd7-en

Tannenbaum, R. J., & Reese, C. M. (2014). *Recommending passing scores for the High School Equivalency Test (HiSET®)* (No. ETS RM-14-06). Educational Testing Service.

Tennessee Department of Labor and Workforce Development. (2024). *New Pathways*. Tennessee State Government. https://www.tn.gov/workforce/jobs-and-education/ae/aer/new-pathways.html

U.S. Census Bureau. (2017). Table 1. Educational attainment of the population 18 years and over, by age, sex, race, and Hispanic origin: 2016 [Dataset]. In *Educational attainment in the United States: 2016*. https://www.census.gov/data/tables/2016/demo/education-attainment/cps-detailed-tables.html

U.S. Census Bureau. (2022). POV-09. Years of school completed by poverty status, sex, age, nativity, and citizenship [Dataset]. In *Poverty status in 2021*. https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pov/pov-09.html

U.S. Census Bureau. (2023). Table 1. Educational attainment of the population 18 years and over, by age, sex, race, and Hispanic origin: 2022 [Dataset]. In *Educational attainment in the United States: 2021*. https://www.census.gov/data/tables/2022/demo/educational-attainment/cps-detailed-tables.html

U.S. Census Bureau. (2024). S1501: Educational attainment [Dataset]. In *2023 American Community Survey 1-Year estimates*. https://data.census.gov/table?q=Educational%20Attainment&y=2023

WestEd. (2015). *An independent study of the alignment between the 2016 ETS HiSET exam items and the College and Career Readiness Standards for Adult Education*. Author.

Yin, R. K. (2017). *Case study research and applications* (6th ed.). SAGE Publications, Inc.

Zinth, J. (2015). *GED, HiSET and TASC Test: A comparison of high school equivalency assessments*. Education Commission of the States. https://www.ecs.org/clearinghouse/01/20/38/12038.pdf

About the Author

Austin S. Jennings
eduCACE
austin@educace.org
https://orcid.org/0000-0002-3342-2205
Austin is the Executive Director of eduCACE and a Senior Research Fellow at DiplomaSender. His research focuses on issues of equity in adult education and data use across education settings and stakeholders.

education policy analysis archives