LEAPS STUDENT VOICE SURVEY

TECHNICAL REPORT

Sanford Student, Susan Lyons, & Samuel D. Ihlenfeldt

Assessment development and validation occurred in close collaboration with the evaluation and measurement team at Transcend: David Nitkin, Youssef Shoukry, Ali Picucci, and Amber Cross

April 2024

Lyons ASSESSMENT CONSULTING

Transcend BUILDING BEYOND THE LIMITS OF SCHOOL DESIGN

Table of Contents

Executive Summary

The purpose of this report is to provide an overview of the technical documentation associated with the design, development, and ongoing validation of the Leaps Student Voice Survey. The Leaps Student Voice Survey was developed for use by Transcend in its redesign work with schools related to the ten Leaps for Equitable, 21st-Century Learning. The Leaps Student Voice Survey consists of a set of eleven psychometric scales:

- Deep Dive Leaps Scales (ten standalone assessments) aimed at providing deeper understanding and insight related to learner experiences for each Leap, and
- Leaps Pulse Check Survey that serves as a quick, diagnostic tool to measure learner experiences related to all ten Leaps.

The current report contains the intended purpose and uses of the Leaps Student Voice Survey as articulated in a theory of action (Figure 1), the design and development processes, and the methods and results of analyses aimed at gathering validity evidence to support the technical quality of the instrumentation. In summary, validity evidence related to the content, cognitive processes, internal structure, reliability, and relationships with other variables indicates strong support for interpreting and using the Leaps Pulse Check and Deep Dive Scales as intended within the Leaps Student Voice Survey System.

The manual is structured into five major sections:

1. Initial development of the Leaps Student Voice Survey,
2. Construction and validation of the Deep Dive Scales,
3. Construction and validation of the Pulse Check Scales,
4. Construction and validation of the Spanish Leaps survey, and
5. Case studies supporting the use of the system.

The scales are freely available for educational and research use. Complete versions of each of the scales are provided in Appendices A and C.

Introduction and Theoretical Framework

In early 2021, Transcend partnered with Lyons Assessment Consulting to develop a measurement system that would support deeper understanding and analysis of learner experiences within its partner schools. Transcend is a national nonprofit that supports school communities to create and spread extraordinary, equitable learning environments. In 2020, Transcend released the framework of the Ten Leaps for Equitable, 21st-Century Learning, summarized in Table 1. The Leaps framework was developed after conducting a thorough and systematic synthesis of the latest research on the science of learning and development, and is now used to guide Transcend's transformation work with schools.

Table 1. Transcend's Ten Leaps for Equitable, 21st-Century Learning

Leap	Description
High Expectations with Unlimited Opportunities	All learners experience high expectations and have equitable access to many opportunities, enabling them to progress toward their aspirations for themselves, their families, and the community—regardless of the time and support needed.
Whole-Child Focus	Learners engage in experiences that nurture the totality of cognitive, emotional, social, and physical factors that impact their learning, development, character, and overall health and well-being.
Rigorous Learning	Learners use critical thinking skills to make deep meaning of diverse, complex ideas and are assessed on their ability to apply, analyze, and use their knowledge in creative ways across contexts.
Relevance	Learning explores young peoples' interests and goals, is connected to their communities, and enables them to understand and tackle real problems in authentic contexts.
Affirmation of Self & Others	Each learner develops a unique, positive sense of self and purpose as well as a deep respect for the identities of others; these diverse identities are celebrated, nurtured, and leveraged in meaningful and anti-oppressive ways to support everyone's learning.

Social Consciousness & Action	Learners critically examine social problems and work toward a more just world; they develop the knowledge, skills, and mindsets needed to continue taking anti-oppressive actions that disrupt and dismantle racism and other inequities.
Connection & Community	The environment is relationship-rich: learners are deeply known and respected by a variety of adults and peers; collaborate closely; and form meaningful relationships across lines of difference that nurture empathy, foster belonging, support well-being, and build social capital.
Customization	The focus, pace, and sequence of learning, as well as the resources and supports provided, are tailored to each learner's identity, prior knowledge, development, way of learning, and life experiences, ensuring that all learners have what they need to be successful and those who need more receive more.
Active Self-Direction	Young people are active drivers of their learning; they grapple directly with concepts while receiving adult and peer guidance and support; they have a voice in decisions about how and what they learn, so that the process grows agency and meaningfully builds on their interests and prior knowledge.
Anytime, Anywhere Learning	Learning can happen anywhere and at any time for all learners with teachers, families, community members, and other important figures in a young person's life, all playing important educational roles.

Lyons Assessment Consulting is a leader in designing innovative assessment systems intended to disrupt systems of oppression and promote social justice. This report details the process and outcomes associated with developing a measurement system that supports Transcend's work with its partner schools using the ten Leaps framework.

Leaps Student Voice Survey Theory of Action

High-quality systems of assessment are designed with a well-articulated theory of action that details how the assessments are intended to work together in service of larger programmatic goals. Figure 1 outlines a high-level theory of action for the Leaps Student Voice Survey. The goal that the measurement system is aiming to support is articulated on the right-most side of the figure, the components of the assessment system are indicated

in dark blue, and the actions and assumptions associated with system use are indicated in light green.

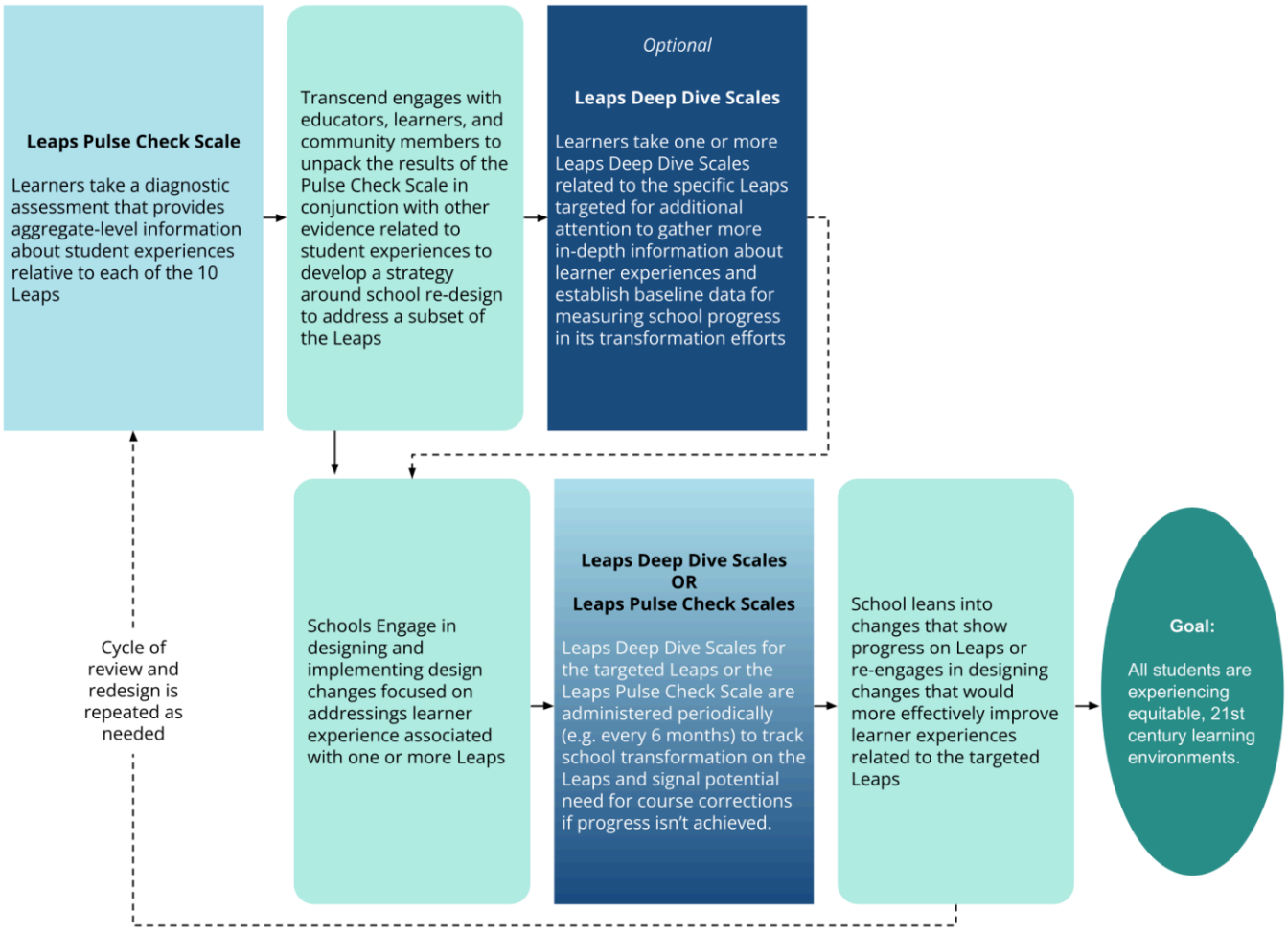


Figure 1. Leaps Student Voice Survey Theory of Action

Overview of Scale Development and Validation

As detailed in Figure 1, the Leaps Student Voice Survey comprises two primary assessment components:

1. **Deep Dive Leaps Scales** (ten standalone assessments) designed to provide a more in-depth measurement of student experiences relative to each Leap.
2. **A Leaps Pulse Check Survey** that captures learner experiences relative to all ten Leaps to provide a quick, diagnostic view of student perceptions and experiences related to all Leaps.

The scale development and validation processes are summarized in Table 2.

Table 2. Summary of Scale Development and Validation

Timeframe	Activity
Winter 2021	Conceptualizing assessment system
	Consulting content expert for each Leap to identify central sub-constructs that comprise the definition of each Leap
Spring 2021	Reviewing literature to identify existing scales related to the sub-constructs represented within each Leap
	Identifying items within existing scales that best represent the intended Leap
	Seeking permission to use and/or modify existing items
	Consulting content experts to review to modify and write new items where needed
	Conducting cognitive labs with students to review and revise newly-developed items
	Piloting the ten Deep Dive Leaps Scales with sample of Transcend's partner schools
Summer 2021	Analyzing pilot data to update Deep Dive Leaps Scales
	Conducting another round of cognitive labs with a subset of scales
	Constructing the Leaps Pulse Check Scale using item statistics and content expert judgment from the Deep Dive Leaps Scales
Fall & Winter 2021-2022	Administering the Leaps Pulse Check Scale with large sample of Transcend's partner schools
	Analyzing administration data and finalizing Leaps Pulse Check Scale
Spring 2022	Writing preliminary technical documentation and recommending next steps for ongoing improvement and technical maintenance of the Leaps Student Voice Survey
Spring/Summer 2022	Collecting additional data for validation work related to Deep Dive Scales
Fall 2022	Analyzing data collected in Spring/Summer 2022

	Administering Spanish Language Leaps Pulse Checks
Spring 2023	Collecting additional data for validation work related to both Deep Dive Scales and Pulse Checks
Fall 2023	Complete set of analysis for all Pulse Checks, and Deep Dives are re-run as final set of validity analyses for current version of scales
Spring 2024	Current version of technical report released

Construction and Validation of Leaps Deep Dive Scales

Summary

This section details the following:

1. Initial development of the Leaps Deep Dive Scales
2. Description of the Spring 2021 Pilot Study
3. Validity analyses for all ten finalized Leaps Deep Dive Scales

Scale Development

Our scale development approach began with a desire to leverage existing validated scales that target similar constructs to those of the Leaps. The intention was to adopt or adapt as many existing items as possible in order to benefit from the existing validation work that went into the development of other scales. Scales were primarily drawn from the academic literature using a library database search, but also were found through other partner non-profits engaged in similar work with schools. All items used or adapted for our scales are either explicitly open-source or were modified with written permission from the scale copyright holders.

In order to find scales that best aligned with the intended measurement targets of our Leaps, Lyons Assessment Consulting worked closely with content experts at Transcend to identify a set of sub-constructs that comprise the definitions of each of the Leaps. Not only was this helpful for our literature search, it also served as an important step in more fully operationalizing the definitions of the Leaps for the purpose of measurement. Table 3 shows the sub-constructs identified for each of the ten Leaps and the source scales from which we were able to adopt or adapt items.

Table 3. Targeted Sub-Constructs that Comprise Each Leap

Leap	Sub-Constructs	Source Scales
High Expectations with Unlimited Opportunities (HE)	High expectations	– Aldridge & Fraser, 2008 – Hart et al., 2021 – MCIEA, 2021
	Equitable access to opportunities	
Whole-Child Focus (WCF)	Support for social-emotional learning	N/A all new items
	Support for physical health	
	Support for mental health	
	Support for identity development	
Rigorous Learning (RL)	Critical thinking	– Ferguson, 2010 – Chai et al., 2015
	Task rigor	
Relevance (REL)	Attention to learner goals/interests	– Assor et al., 2002 – Burns, 2006 – Chai et al., 2015 – Frymier & Shulman, 1995 – Young et al., 2008
	Connection to the real world	
	Connection to prior learning	
Affirmation of Self & Others (ASO)	Affirmation in school	– MCIEA, 2021
	Diversity and inclusion	
Social Consciousness & Action (SCA)	Support for critical consciousness	– Panorama, 2021
Connection & Community (CC)	Connections to adults	– Appleton et al., 2006 – Chai et al., 2015 – MCIEA, 2021
	Connections to peers	
	Connection to community	
	Collaboration	
Customization (CUS)	Personalized support	– Aldridge & Fraser, 2008 – Assor et al., 2002 – Ferguson, 2010 – Hart et al., 2021
	Personalized pace	
	Personalized materials and resources	

		– US Department of Education, 2017
Active Self-Direction (ASD)	Student choice	– Appleton et al., 2006 – Assor et al., 2002 – Burns, 2006 – Ferguson, 2010
	Empowerment at school	
	Self-directed learning	
Anytime, Anywhere Learning (AAL)	Learning outside of school building	N/A all new items
	Learning outside of school hours	
	Valuing other sources of learning	

Even with the excellent survey sources from which we drew our initial set of items, we nonetheless needed to engage in some new item development for all scales. Item development occurred iteratively with rounds of input, review, and revision from the Lyons Assessment Consulting and Transcend teams. We relied heavily on the content expertise of the Transcend team who were closely involved in the Leaps framework development. Additionally, we drew from existing resources in the field to help shape our understanding about the most central ideas associated with each Leap. For example, to help us define the sub-constructs and ultimately write items associated with the Whole Child Focus, we drew on the Whole Child Framework developed by the Chan-Zuckerberg Initiative (CZI); see Figure 2.

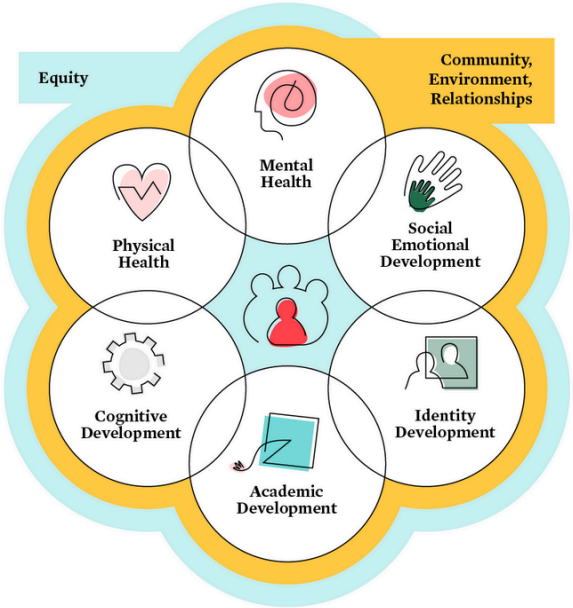


Figure 2. Chan-Zuckerberg Initiative Whole Child Framework

Because multiple Leaps attend to cognitive and academic development, item development for the Whole Child Focus Leap focused on the physical health, mental health, social-emotional development, and identity development aspects of the CZI framework.

All items that were newly developed for our Spring 2021 piloting of the Leaps Deep Dive Scales were tested with students using a cognitive laboratory protocol. The purpose of the cognitive laboratory is to gauge the degree to which the items, as written, are eliciting the intended cognitive processes as students read and respond to the scales. Seven students ranging from grades 3–9 participated in these cognitive labs in which they read the item aloud and then shared their thinking as they selected their response to the item on the provided Likert scale. In some cases, item wording was difficult to understand or elicited thinking that was not in line with the intended meaning of the item. These cognitive labs were thus highly informative for making item revisions and refinements before we conducted formal pilot testing with a large sample of students.

Piloting and Scale Revisions

Following initial development, each Deep Dive Scale was piloted during the Spring of 2021. Students across six schools participated in the pilot, with a total of 1,670 students responding to at least one Leap scale. The majority were in grades 3–8, with a small number from one school in grade 1.

School leaders chose to administer one or more of the Leap Deep Dive Scales to their students.

The following analyses were conducted to make scale recommendations for the finalized Leaps Deep Dive Scales. For each of these analyses, a greater description is offered in the "Scale Analysis" section below, where we describe them in the context of the current validation work:

- Mean responses and standard deviations (SDs) were calculated for all items individually, each scale overall, and disaggregated by student group.
- For each scale, item responses were analyzed using exploratory factor analysis (EFA). EFA is a method of analyzing the covariance of items on a given scale to understand the extent to which they appear to reflect the same underlying construct(s).
- A reliability analysis was conducted in order to inform the reduction of each scale.
- For each item, differential item functioning (DIF) analysis looks for evidence that respondents of similar overall levels on the underlying construct differ

systematically in their responses to the given item according to known demographic variables.

The primary purpose of the pilot was to identify potential items for removal from each scale, as the original scales consisted of approximately 15 items each. The target final scale length was between 8–10 items. These initial analyses were intended to flag any items unlikely to be useful due either to extreme response patterns (e.g., most students selecting the highest or lowest level of agreement) or to potentially targeting a different construct (as indicated by a low item-total correlation). In many cases, reducing the scale was feasible, and a number of items were flagged and removed.

Lastly, during piloting, the opportunity arose to pilot a couple of the Deep Dive Scales with students in first grade. The intent here was to explore the potential appropriateness for using these scales with students under the intended grade 3 reading level required to complete the Leaps Deep Dive Scales. These analyses indicated that the responses gathered from grade 1 students appeared to be less trustworthy and reliable compared to those gathered from older students. As such, we do not currently recommend the Leaps Student Voice Survey for use with students below grade 3.

Current Validation Work

Administration Summary

The current versions of the Leaps Deep Dive Scales are based on recommendations from analyzing the previous administrations of earlier versions of each scale. The full operational text for each item can be found in Appendix A.

The data used for the most recent analyses were largely collected during the Fall of 2022 and the Spring of 2023, however, for some scales (HE, CC) the items did not change at all from the previous version, so we have included responses from prior data collection periods. Demographic breakdowns of the samples can be found by grade (Table 4) and by race and ethnicity, gender, and free and reduced-price lunch status (FRL; Table 5). As a note, because of our large sample size, we have subsetted the data to just respondents who completed all items in each scale, respectively. Some of the scales (SCA, CUS, ASD, AAL, and to a lesser extent, RL and REL), were not administered fully as Deep Dives to students in grades 3–5. Consequently, for those scales, the present analyses cannot be extrapolated below grade 6 and we caution against their use in those grades unless more data is collected.

Table 4. Number of Full Responses by Scale and Grade

Scale	N	Gr 3	G 4	Gr 5	Gr 6	Gr 7	Gr 8	Gr 9	Gr 10	Gr 11	Gr 12
HE	3,007	174	185	191	312	601	749	269	243	176	107
WCF	4,339	895	926	850	175	370	439	207	187	182	108
RL	1,925	16	9	0	244	459	513	173	185	169	157
REL	1,330	26	14	10	91	273	427	131	132	119	107
ASO	4,357	904	940	858	177	366	434	210	186	176	106
SCA	1,116	0	0	0	61	253	315	131	131	118	107
CC	5,427	1,069	1,082	965	351	488	671	270	245	180	106
CUS	1,558	0	0	0	201	414	466	130	127	110	110
ASD	1,815	0	0	0	241	426	661	131	131	117	108
AAL	1,135	0	0	0	57	248	353	133	129	112	103

Table 5. Number of Full Responses by Demographic Categories

Scale	Race/Ethnicity							Sex			FRL Status		
	Afr. Amer. or Black	Amer. Indian or Alaska Native	Asian or Pac. Isl.	Hisp. or Latino	Two or More Races	White	Missing	Female	Male	Missing	No	Yes	Missing
HE	666	10	79	1,566	71	568	47	1,530	1,477	0	338	1,871	798
WCF	761	10	70	2,632	50	184	632	2,039	2,092	208	328	813	3198
RL	688	1	86	804	105	241	0	929	996	0	353	1,572	0
REL	237	27	61	788	56	152	9	630	700	0	385	881	64
ASO	777	12	69	2,635	48	185	631	2,054	2,087	216	313	808	3236
SCA	216	1	48	765	27	59	0	511	605	0	309	807	0
CC	941	41	88	2,977	72	622	686	2,639	2,572	216	396	1,665	3366
CUS	419	2	82	765	80	210	0	718	840	0	310	1,248	0
ASD	413	1	268	861	80	192	0	860	955	0	349	1,438	28
AAL	218	1	51	755	34	76	0	523	612	0	325	810	0

Item & Scale Statistics

In this section, for all scales, we provide classical item statistics including mean, standard deviation, and item-total correlation (Table 6); Cronbach's alpha lower-bound estimates of reliability (Table 7); and results from exploratory factor analysis. Figure 3 depicts the distribution of mean responses for each of the ten Deep Dive Scales.

HE

The total number of complete responses to the HE Deep Dive Scale was 3,007. Both Cronbach's α and Guttman's λ_6 were .91, which indicates a very high degree of internal consistency in the scale. The mean response out of five was 3.35 with a standard deviation of 0.74. Classical item statistics indicate high levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

WCF

A total of 4,339 complete responses were collected for the WCF Deep Dive Scale. Both Cronbach's α and Guttman's λ_6 were .92, which indicates a very high degree of internal consistency in the scale. The mean response out of five was 3.50 with a standard deviation of 0.84. Classical item statistics indicate moderately high levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

RL

There were 1,925 complete responses to the RL Deep Dive Scale. Both Cronbach's α and Guttman's λ_6 were .92, which indicates a very high degree of internal consistency in the scale. The mean response out of five was 3.28 with a standard deviation of 0.75. Classical item statistics indicate high levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

REL

We collected 1,330 complete responses to the REL Deep Dive Scale. Cronbach's α was .94 and Guttman's λ_6 was .93, which both indicate a very high degree of internal consistency in the scale. The mean response out of five was 2.89 with a standard deviation of 0.90. Classical item statistics indicate moderately high levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

ASO

The total number of complete responses to the ASO Deep Dive Scale was 4,357. Both Cronbach's α and Guttman's λ_6 were .89, which indicates a high degree of

internalconsistency in the scale. The mean response out of five was 3.52 with a standard deviation of 0.81. Classical item statistics indicate moderately high levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

SCA

A total of 1,116 complete responses were collected for the SCA Deep Dive Scale. Both Cronbach's α and Guttman's λ_6 were .91, which indicates a very high degree of internal consistency in the scale. The mean response out of five was 2.80 with a standard deviation of 0.80. Classical item statistics indicate moderately high levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

CC

There were 5,427 complete responses to the CC Deep Dive Scale. Both Cronbach's α and Guttman's λ_6 were .91, which indicates a very high degree of internal consistency in the scale. The mean response out of five was 3.51 with a standard deviation of 0.84. Classical item statistics indicate moderately high levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

CUS

We collected 1,558 complete responses to the CUS Deep Dive Scale. Both Cronbach's α and Guttman's λ_6 were .94, which indicates a very high degree of internal consistency in the scale. The mean response out of five was 3.18 with a standard deviation of 0.76. Classical item statistics indicate high levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

ASD

The total number of complete responses to the ASD Deep Dive Scale was 1,815. Cronbach's α was .90 and Guttman's λ_6 was .89, which both indicate a high degree of internal consistency in the scale. The mean response out of five was 2.96 with a standard deviation of 0.84. Classical item statistics indicate moderately high levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

AAL

A total of 1,135 complete responses were collected for the AAL Deep Dive Scale. Both Cronbach's α and Guttman's λ_6 were .93, which indicates a very high degree of internal consistency in the scale. The mean response out of five was 3.04 with a standard deviation of 0.72. Classical item statistics indicate moderately high levels of agreement with variability

among individual respondents, and item-total correlations are high. However, two items, “My school provides technology to support my learning.” (V2_AAL_Q_7) and “I am able to access my school online.” (V2_AAL_Q_13), showed relatively lower correlations. As a result, the final survey instruments will exclude V2_AAL_Q_7.

Table 6. Final Set of Items Administered

Scale	Item	M	SD	Item-Total Correlation
HE	V2_HE_Q_1	3.48	1.00	0.65
	V2_HE_Q_2	2.99	1.07	0.56
	V2_HE_Q_3	3.36	0.99	0.67
	V2_HE_Q_6	3.65	0.99	0.69
	V2_HE_Q_7	3.37	1.08	0.69
	V2_HE_Q_8	3.36	0.97	0.7
	V2_HE_Q_9	3.23	1.07	0.68
	V2_HE_Q_10	3.36	0.99	0.75
	V2_HE_Q_11	3.48	0.91	0.67
	V2_HE_Q_12	3.53	0.92	0.67
	V2_HE_Q_13	3.07	1.11	0.59
WCF	V2_WCF_Q_1	3.46	1.17	0.73
	V2_WCF_Q_2	3.45	1.15	0.72
	V2_WCF_Q_3	3.61	1.11	0.73
	V2_WCF_Q_4	3.64	1.1	0.7
	V2_WCF_Q_5	3.32	1.21	0.65
	V2_WCF_Q_6	3.69	1.04	0.64
	V2_WCF_Q_7	3.19	1.19	0.68
	V2_WCF_Q_8	3.65	1.01	0.64
	V2_WCF_Q_9	3.59	1.06	0.69
	V2_WCF_Q_10	3.36	1.2	0.67
	V2_WCF_Q_11	3.55	1.14	0.69
RL	V2_RL_Q_1	3.27	1.02	0.7
	V2_RL_Q_2	3.04	1.04	0.67
	V2_RL_Q_3	3.39	0.97	0.69
	V2_RL_Q_4	3.3	0.98	0.66
	V2_RL_Q_5	3.35	1.07	0.65
	V2_RL_Q_6	3.48	1	0.71

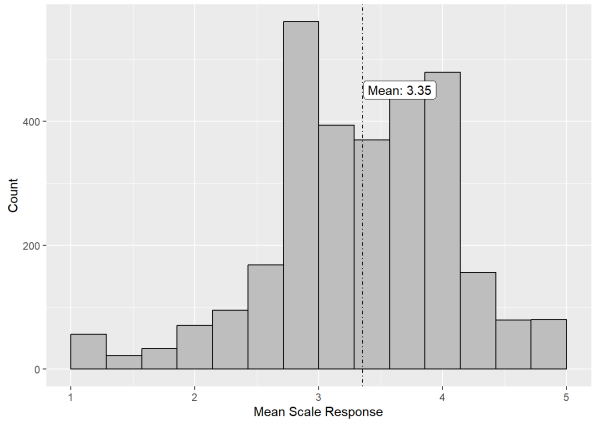
	V2_RL_Q_7	3.24	0.99	0.72
	V2_RL_Q_8	3.44	0.96	0.72
	V2_RL_Q_9	3.48	0.93	0.65
	V2_RL_Q_10	2.99	1.05	0.68
	V2_RL_Q_11	3.1	1.07	0.7
REL	V2_REL_Q_1	2.78	1.2	0.75
	V2_REL_Q_2	2.81	1.15	0.72
	V2_REL_Q_3	3.06	1.18	0.67
	V2_REL_Q_4	2.8	1.18	0.78
	V2_REL_Q_5	2.95	1.12	0.74
	V2_REL_Q_6	2.92	1.13	0.76
	V2_REL_Q_7	2.73	1.14	0.76
	V2_REL_Q_8	2.93	1.1	0.71
	V2_REL_Q_9	2.85	1.11	0.76
	V2_REL_Q_10	3.03	1.1	0.69
	V2_REL_Q_11	2.86	1.16	0.74
ASO	V2_ASO_Q_1	3.27	1.16	0.61
	V2_ASO_Q_2	3.69	1.16	0.68
	V2_ASO_Q_3	3.58	1.2	0.63
	V2_ASO_Q_4	3.6	1.15	0.72
	V2_ASO_Q_5	3.38	1.12	0.58
	V2_ASO_Q_6	3.56	1.04	0.63
	V2_ASO_Q_7	3.6	1.06	0.62
	V2_ASO_Q_8	3.56	1.12	0.62
	V2_ASO_Q_10	3.49	1.09	0.67
	V2_ASO_Q_11	3.47	1.16	0.61
SCA	V2_SCA_Q_2	2.67	1.16	0.72
	V2_SCA_Q_3	3	1.08	0.62
	V2_SCA_Q_4	2.84	1.09	0.61
	V2_SCA_Q_5	2.74	1.09	0.72
	V2_SCA_Q_6	2.52	1.1	0.69
	V2_SCA_Q_7	2.71	1.1	0.68
	V2_SCA_Q_8	2.72	1.15	0.67
	V2_SCA_Q_9	2.85	1.15	0.56
	V2_SCA_Q_13	2.55	1.16	0.7

	V2_SCA_Q_16	3.25	1.09	0.53
	V2_SCA_Q_17	2.91	1.12	0.65
CC	V2_CC_Q_1	3.54	1.07	0.72
	V2_CC_Q_2	3.56	1.22	0.72
	V2_CC_Q_3	3.72	1.06	0.61
	V2_CC_Q_4	3.44	1.06	0.69
	V2_CC_Q_5	3.47	1.08	0.68
	V2_CC_Q_6	3.52	1.13	0.72
	V2_CC_Q_7	3.71	1.02	0.66
	V2_CC_Q_8	3.33	1.12	0.64
	V2_CC_Q_9	3.38	1.19	0.7
	V2_CC_Q_10	3.39	1.16	0.69
CUS	V2_CUS_Q_1	3.05	1.13	0.66
	V2_CUS_Q_2	3.28	0.96	0.71
	V2_CUS_Q_3	3.01	1.07	0.74
	V2_CUS_Q_4	3.08	1.06	0.72
	V2_CUS_Q_5	3.16	1.1	0.7
	V2_CUS_Q_6	3.17	1.12	0.38
	V2_CUS_Q_7	3.06	1.07	0.68
	V2_CUS_Q_8	3.31	1.01	0.71
	V2_CUS_Q_9	3.13	1.05	0.73
	V2_CUS_Q_10	3.31	0.99	0.71
	V2_CUS_Q_11	3.17	1	0.7
	V2_CUS_Q_12	3.03	1.06	0.75
	V2_CUS_Q_13	3.4	1.04	0.67
	V2_CUS_Q_14	3.3	0.95	0.72
	V2_CUS_Q_15	3.31	0.98	0.7
ASD	V2_ASD_Q_1	3.39	1	0.61
	V2_ASD_Q_2	2.68	1.23	0.59
	V2_ASD_Q_3	2.93	1.15	0.67
	V2_ASD_Q_4	3.02	1.1	0.68
	V2_ASD_Q_5	2.76	1.17	0.66
	V2_ASD_Q_6	3.08	1.13	0.68
	V2_ASD_Q_7	2.72	1.17	0.63
	V2_ASD_Q_8	3	1.15	0.71

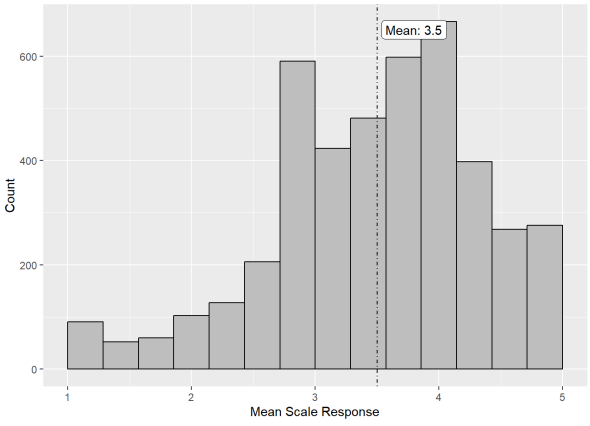
	V2_ASD_Q_9	3.01	1.11	0.7
AAL	V2_AAL_Q_1	2.55	1.14	0.67
	V2_AAL_Q_2	2.82	1.14	0.63
	V2_AAL_Q_3	2.87	1.22	0.53
	V2_AAL_Q_4	3.01	1	0.67
	V2_AAL_Q_5	2.77	1.09	0.69
	V2_AAL_Q_6	3.15	1.09	0.67
	V2_AAL_Q_7	3.72	0.93	0.45
	V2_AAL_Q_8	3.27	1.02	0.56
	V2_AAL_Q_9	2.7	1.1	0.68
	V2_AAL_Q_10	2.9	1.05	0.73
	V2_AAL_Q_11	2.8	1.1	0.7
	V2_AAL_Q_12	2.73	1.09	0.69
	V2_AAL_Q_13	3.68	0.93	0.44
	V2_AAL_Q_14	3.12	1.13	0.59
	V2_AAL_Q_15	3.21	1.05	0.6
	V2_AAL_Q_16	3.21	1.09	0.57
	V2_AAL_Q_17	3.1	1.02	0.71

Figure 3. Mean Response Distributions for Each Deep Dive Subscale

HE

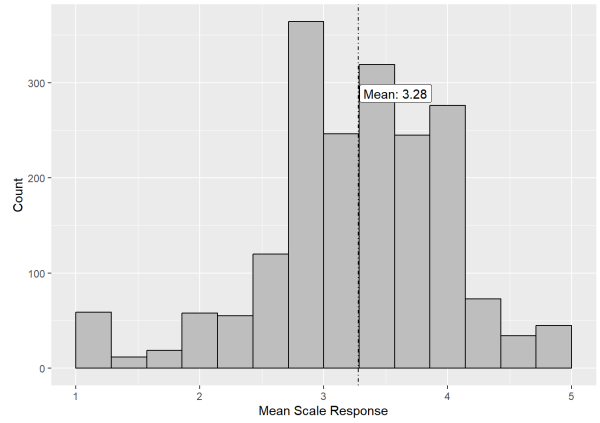


WCF

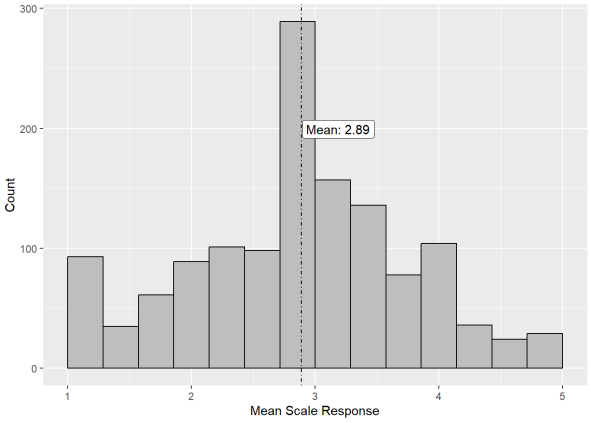


RL

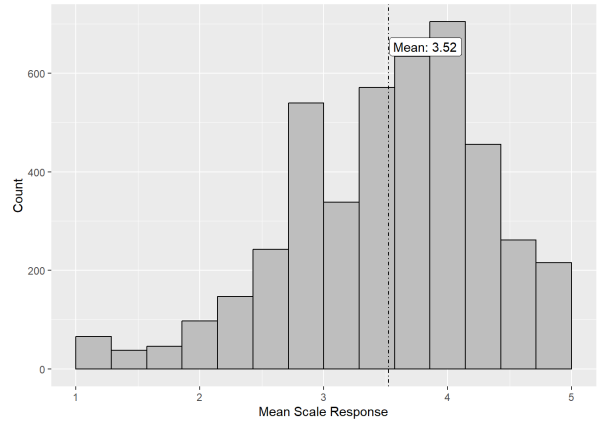
REL



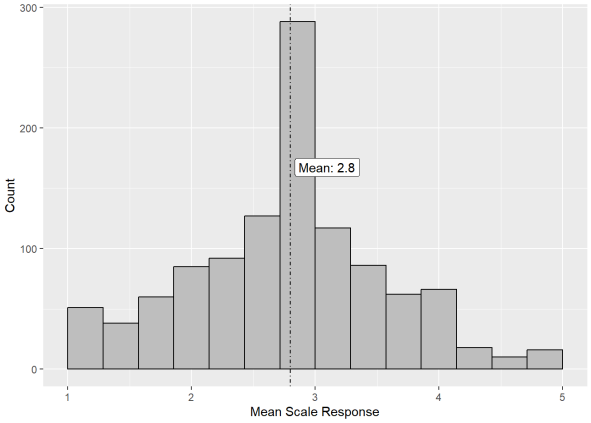
ASO



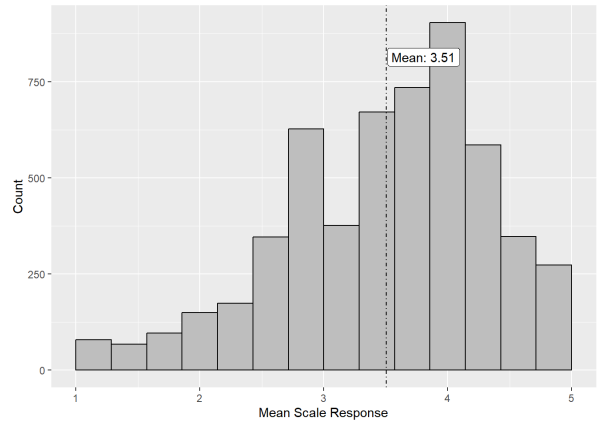
SCA



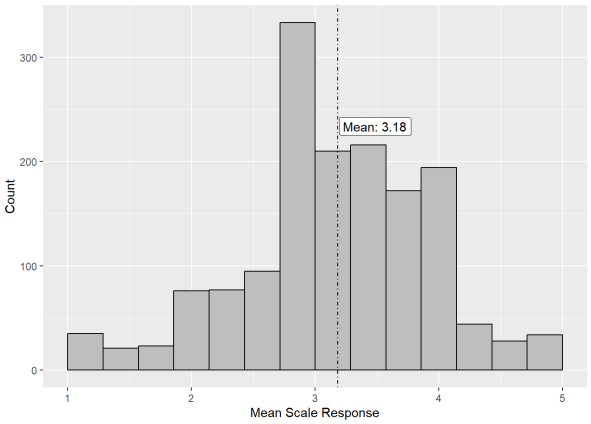
CC



CUS



AAL



ASD

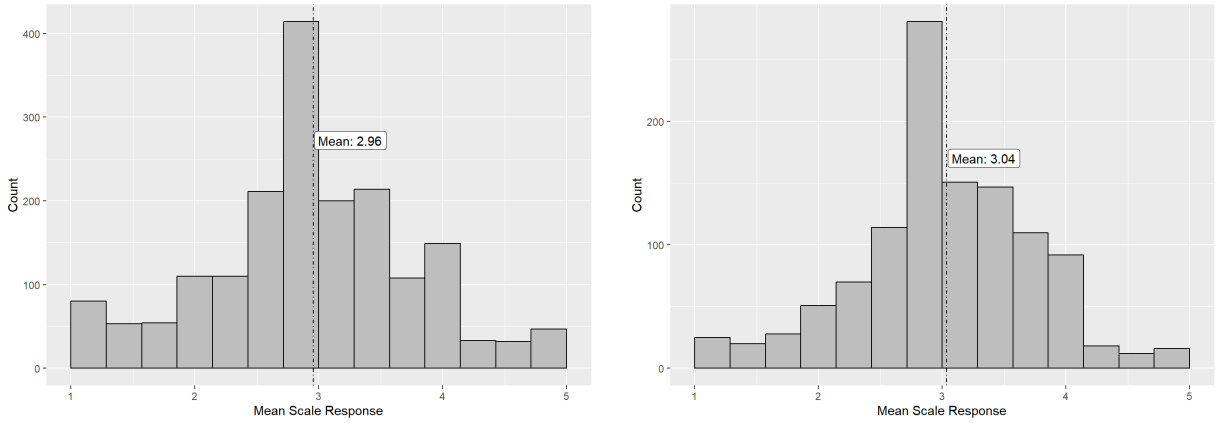


Table 7. Reliability by Deep Dive Subscale

Scale	Cronbach's α	Guttman's λ_6
HE	.91	.91
WCF	.92	.92
RL	.92	.92
REL	.94	.93
ASO	.89	.89
SCA	.91	.91
CC	.91	.91
CUS	.94	.94
ASD	.90	.89
AAL	.93	.93

Exploratory Factor Analysis

For each scale, item responses were analyzed using exploratory factor analysis (EFA). EFA is a method of analyzing the covariance of items on a given scale to understand the extent to which they appear to reflect the same underlying construct(s). Our analysis began with a parallel analysis and only explored multi-factor solutions if the parallel analysis suggested the presence of more than one factor. In these cases, we inspected both RMSR and factor structure for multidimensional solutions to assess whether adding more factors resulted in an interpretable factor structure, or if the unidimensional solution was sufficient.

First, we used visual interpretations of scree tests to inform our factor analysis (Figure 4; Cattell, 1966). A scree plot displays the eigenvalues of factors in descending order against their respective factor numbers. Eigenvalues represent the variance explained by each factor. The plot appears as a line graph, and the point where the line sharply drops off

helps determine the number of factors to retain. The Kaiser-Guttman test (Kaiser, 1960) suggests that factors with eigenvalues greater than 1.0 are considered substantial and typically retained for interpretation. For each Deep Dive Scale, we can see that the first factor is the only one to fall above the 1.0 threshold, which strongly supports a unidimensional factor structure.

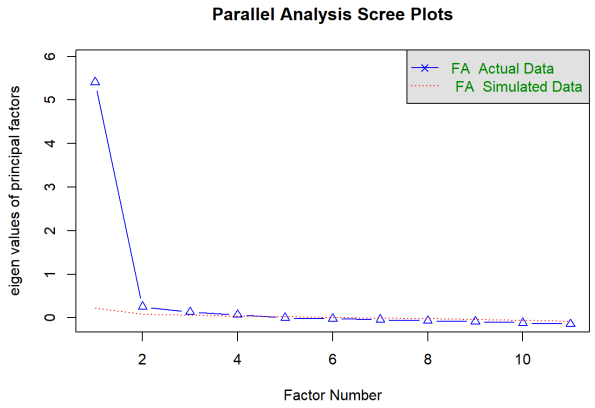
Next, we consider the parallel analysis. Using a simulated data set that matched the size and structure of our actual data, we created a line representing the average eigenvalues obtained from the simulated datasets. It serves as a reference line: if the real eigenvalues surpass the simulated ones, then those factors might be significant and should be retained. Turning to Figure 4, we can see that the parallel analysis suggested the presence of more than one factor for every Deep Dive Scale.

Thus, we turned to the model fit indices and factor structures to explore multifactor solutions, specifically, root-mean-square residuals (RMSR; Table 8). RMSR is a statistical measure used to assess the goodness of fit in structural equation models. It represents the average difference between observed and predicted values, providing an indication of how well the model fits the observed data. In the case of each Deep Dive Scale, RMSR for the one-factor solution fell below .08, which is generally accepted as a good fit (Hornsby et al., 2021) and adding more factors did not produce simpler structures (in which each item loads on a single factor) or any clear breakdown between items from a content perspective.

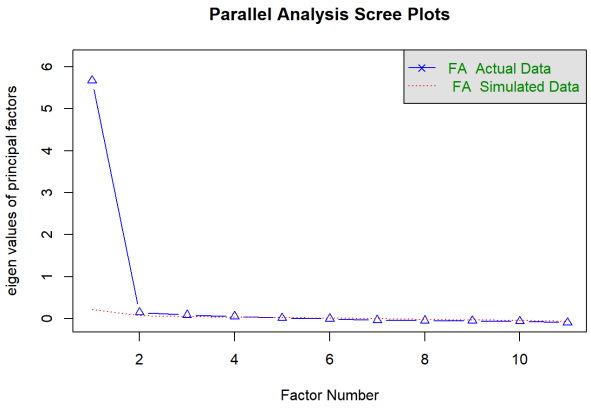
Interpreting a test where items load onto multiple factors can be complex for the end user (Calderón Garrido et al., 2019). When items show cross-loadings across multiple factors, it suggests that those items are associated with more than one underlying construct or dimension in the dataset. The complexity introduced by items loading across various factors makes it harder to simplify and communicate the test's outcomes in a clear and straightforward manner, hampering its practical usability for end users. However, goodness of fit indices are only one consideration when determining dimensionality (Calderón Garrido et al., 2019), and a unidimensional solution is otherwise supported.. Consequently, we support the unidimensional interpretation of all ten Leaps Deep Dive Scales for the purposes of summarizing across item responses.

Figure 4. Scree Plots for Each Deep Dive Subscale

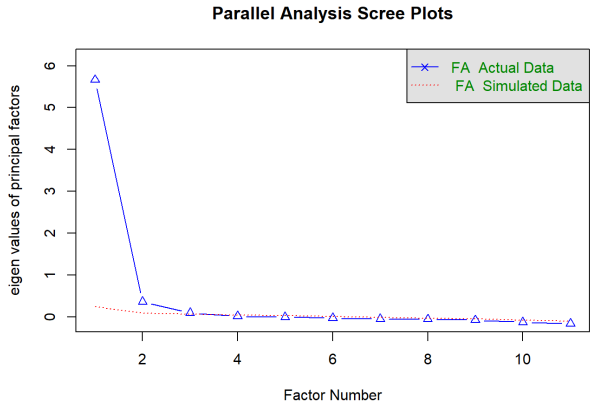
HE WCF



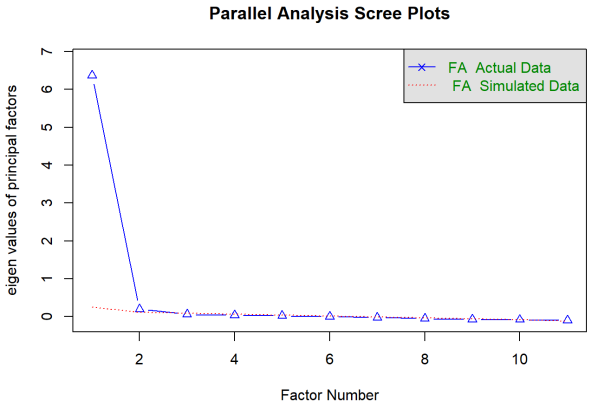
RL



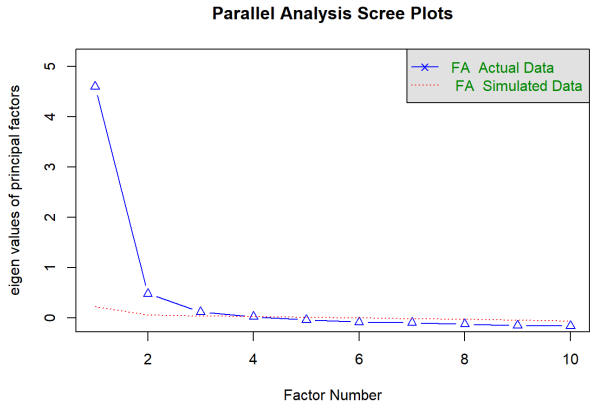
REL



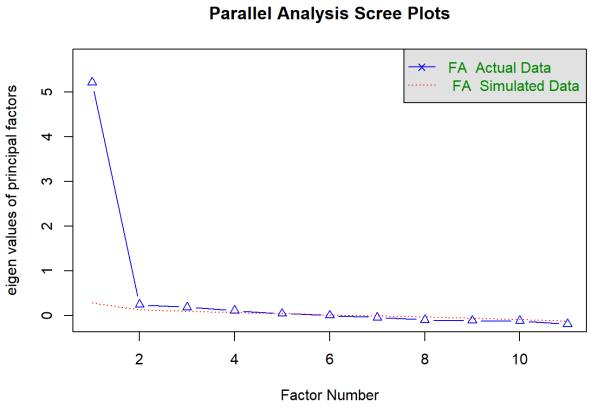
ASO



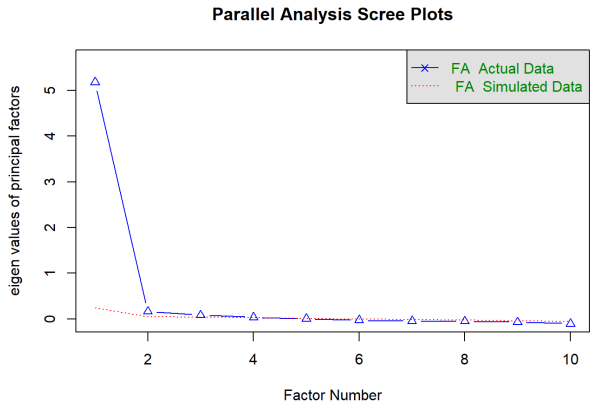
SCA



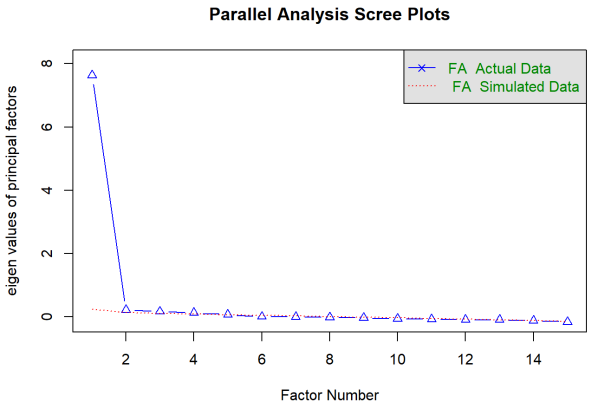
CC



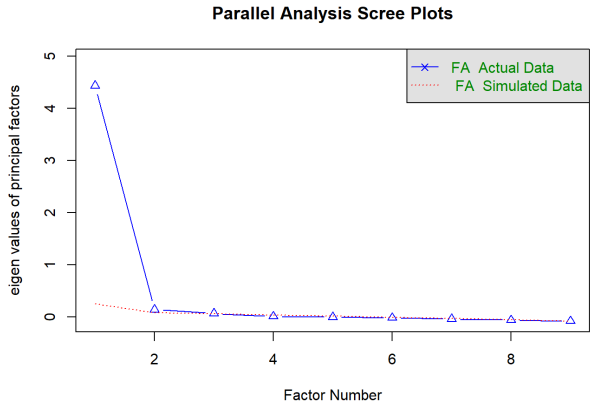
CUS



ASD



AAL



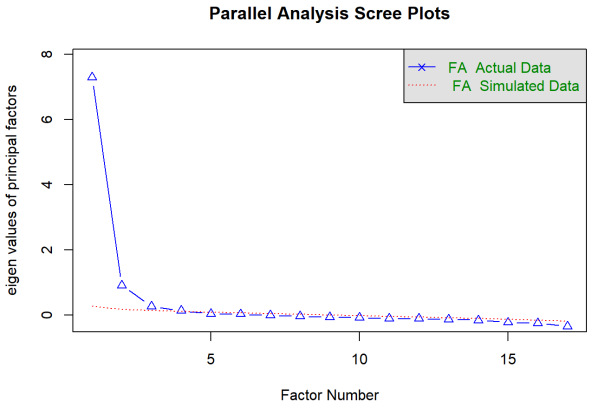


Table 8. Unidimensional RMSR Values for Each Deep Dive

Scale	RMSR
HE	0.035
WCF	0.021
RL	0.042
REL	0.025
ASO	0.060
SCA	0.041
CC	0.025
CUS	0.028
ASD	0.022
AAL	0.067

DIF Results

Our analysis of measurement invariance for each scale focused primarily on differential item functioning (DIF). For each item, DIF analysis looks for evidence that respondents of similar overall levels on the underlying construct differ systematically in their responses to the given item according to known demographic variables. An example of DIF would be a situation in which Black students with similar overall scores to White students on the WCF scale tended to respond substantially lower to a specific item on the scale. If the difference is large enough, this situation would have the potential to attenuate group-level summaries of responses not only to the item itself, but to the scale as a whole, potentially misleading stakeholders attempting to improve whole-child strategies in their classroom or school.

There are many methods for detecting DIF; we selected logistic ordinal regression DIF analysis. Logistic regression DIF analysis was selected as the DIF detection method for this analysis because it is easily implemented for polytomous item responses such as ours, as outlined in Zumbo (1999). Logistic ordinal regression DIF analysis looks for DIF by fitting a logistic regression model where the outcome variable is correctly responding to an item and where the predictor variables are group membership, sum score across all items of the scale, and an interaction of the two.

Logistic regression DIF analysis consists of up to three regression models per item. The first regresses responses to a given item on examinees' total score. The second adds the grouping variable—for example, gender—to the analysis. The third adds an interaction term to the model. A chi-square test of the third model against the first with two degrees of freedom then serves as a simultaneous test of uniform and non-uniform DIF. One can also test the second model against the first as a test of just uniform DIF. Zumbo (1999) contains a great deal of information on this method. For this analysis, our focus was on the simultaneous test of uniform and non-uniform DIF, which provides the most information on the presence of any DIF for a given item. An item was flagged for statistically significant DIF if the chi-squared test of differences produced a p-value less than 0.05. This is a very generous p-value considering the number of hypothesis tests conducted; it was selected out of an abundance of caution to ensure that even borderline DIF could be detected.

Statistical significance is only part of the story in logistic regression DIF analysis; one must then calculate DIF "effect sizes" to help decide if the DIF is meaningful or not. To do this, one must look at the magnitude of DIF for all items (per Zumbo's advice to calculate all effect sizes, then make sense of them in tandem with statistical significance). In logistic

regression DIF analysis, this is represented by the difference in R-squared between models 1 (item response on total score) and 3 (item response on total score, group membership and interaction). Again, in keeping with Zumbo (1999), we calculated the McKelvey-Zavoina pseudo-R-squared measure for each model and differentiate between small and large amounts of DIF with a cutoff of 0.13. This cutoff can be stratified further; we treated a difference of less than 0.05 as trivial.

For most scales, the group with the largest sample size in the item response dataset was treated as the reference group—for our dataset, this was Hispanic/Latino students for every scale. Analyses were conducted for both gender and race/ethnicity. All analyzed scales had sufficient sample size for analysis of potential DIF by gender. For race/ethnicity, analyses were generally only conducted for pairs of race/ethnicity variables with more than 60 students. Thus, based on the available demographic information (Table 5), the following comparisons were made for each of the ten Leaps Deep Dive Scales:

- Female and Male
- Hispanic/Latino and Black
- Hispanic/Latino and White
- Hispanic/Latino and Asian/Pacific Islander

The full results of the DIF analysis can be found in Appendix B. Across all comparisons, our analysis identified a small handful of items showing statistically significant DIF. However, as noted previously, the focus of a DIF analysis needs to go beyond statistical significance. Thus, we opted to use the very conservative alpha value of $p < 0.05$, which was not corrected for the large number of comparisons run in this set of analyses. A p-value correction that is too stringent could result in erroneously ignoring statistically significant DIF that would otherwise be practically significant; instead, for the handful of items which were identified as displaying statistically significant DIF, it was necessary to consider the practical significance of the effect sizes. Doing so, we found that none of the items showing statistically significant DIF displayed anything beyond a trivial effect size. As a result, we conclude that statistically significant DIF is not a salient issue for any of the Deep Dive Scale items.

Construction and Validation of Leaps Pulse Check

Summary

This section details the following:

4. Initial development of the Leaps Pulse Check Scales

5. Description of the Fall 2021 Pilot Study
6. Validity analyses for the Leaps Pulse Check Scales

Scale Development

During the development of the ten Deep Dive Leaps Scales, the Transcend team recognized that it could be valuable if a school had access to a single brief survey form that could support diagnostic interpretations at the school-level about learning experiences on all ten Leaps—a Leaps Pulse Check Scale. The Pulse Check Scale development requirements were to (a) align to our theory of action, (b) be sufficiently reliable to support analyses related to the Leaps aggregated at the classroom-level, and (c) contain enough items to support a separate interpretation for each Leap. Additionally, the Pulse Check Scale would ideally take students a fairly short amount of time to complete.

The Pulse Check Scale contains items from all of the ten Leaps scales. The initial design of the Pulse Check Scale required that for every Leap, enough items were included to achieve at least a coefficient α of 0.70. The Pulse Check Scale was the result of making informed decisions about item selection in a way that balanced the known psychometric properties of the items (e.g., response distributions, item-total correlations, estimated subscale reliably), as well as equally-important content-based decisions related to construct representation.

Piloting and Scale Revisions

Item responses were collected during Fall 2021. Participating schools came from two projects in which Transcend is engaged: one with rural schools, and one with schools in Texas. In total, 7,662 students from grades 3–12 answered a version of the Leaps Pulse Check survey.

The following analyses were conducted to make scale recommendations for the finalized Leaps Pulse Check Scales:

- Mean responses and standard deviations (SDs) were calculated for all items individually, each scale overall, and disaggregated by student group.
- A reliability analysis was conducted in order to inform the reduction of each scale.
- To analyze internal structure, the entire sample was divided in half. For the first half of participants, item responses were analyzed using exploratory factor analysis (EFA). In most cases, the short length of the subscale led to negative degrees of freedom for a solution with two or more factors. As a result, we focused mainly on the fit of the unidimensional model.

- Once an internal structure was identified through EFA, a secondary confirmatory factor analysis (CFA) was performed which confirmed the hypothesized factor structure of the entire Pulse Check Scale comprising the ten underlying subscales.
- Due to high correlations between the Leaps Pulse Check Scales, a Principal Components Analysis (PCA) was conducted in order to determine if a single overall score measuring school progress could be derived from averaging across the 10 subscales.
- We conducted analyses of variance (ANOVAs) for all Leaps. These were intended to provide insight into the extent to which variance in students' mean responses to each Leap could be accounted for using a combination of school identifiers and student demographics. In all cases, the school that a student attends accounted for a statistically significant amount of variance in scale responses. We also noted that gender was found to account for significant variance on several scales.

At the conclusion of the pilot phase of the Pulse Check Scale, we recommended the inclusion of a number of field test items for the next round of survey administration.

Current Validation Work

Administration Summary

Like the Deep Dive Scale data, the data were collected during the Fall of 2022 and the Spring of 2023. Demographic breakdowns of the samples can be found by grade (Table 9) and by ethnicity, gender, and FRL status (Table 10).

Table 9. Number of Full Responses by Scale and Grade

Scale	N	Gr 3	Gr 4	Gr 5	Gr 6	Gr 7	Gr 8	Gr 9	Gr 10	Gr 11	Gr 12
HE	40,602	1,704	2,125	3,131	6,686	6,818	7,356	3,941	3,622	3,113	2,106
WCF	29,009	1,778	2,094	2,794	4,362	4,338	4,516	2,851	2,522	2,210	1,544
RL	41,938	1,798	2,214	3,261	6,982	6,993	7,596	4,008	3,721	3,186	2,179
REL	40,435	1,733	1,933	3,006	6,650	6,780	7,443	3,973	3,666	3,131	2,120

ASO	29,220	2,028	2,113	3,086	4,399	4,335	4,596	2,691	2,422	2,105	1,445
SCA	37,744	1,426	1,580	2,395	6,202	6,514	6,835	3,938	3,645	3,101	2,108
CC	30,788	2,137	2,439	3,368	4,710	4,444	4,901	2,761	2,476	2,106	1,446
CUS	41,240	1,714	2,147	3,173	6,928	6,957	7,450	3,961	3,653	3,130	2,127
ASD	26,705	1,000	1,098	2,157	4,485	4,443	4,941	2,661	2,393	2,089	1,438
AAL	27,458	1,082	1,181	2,250	4,545	4,515	4,902	2,773	2,503	2,184	1,523

Table 10. Number of Full Responses by Demographic Categories

Scale	Race/Ethnicity							Sex			FRL Status		
	Afr. Amer. or Black	Amer. Indian or Alaska Native	Asian or Pacific Isl.	Hisp. or Latino	Two or More Races	White	Missing	Female	Male	Missing	No	Yes	Missing
HE	12,583	175	2,203	12,183	2,284	9,961	943	20,623	19,920	59	3,607	32,445	4,550
WCF	9,781	83	1,555	8,354	1,787	5,887	1,562	14,448	14,281	280	2,308	22,406	4,295
RL	13,438	177	2,268	12,497	2,358	10,237	963	21,234	20,642	59	3,682	33,657	4,599
REL	12,784	204	2,223	12,083	2,275	9,895	971	20,522	19,857	56	3,542	32,253	4,640
ASO	8,440	101	1,556	9,718	1,808	6,031	1,566	14,537	14,402	281	2,344	20,846	6,030
SCA	12,663	174	2,222	10,311	2,244	9,166	964	19,128	18,557	59	3,004	31,833	2,907
CC	8,790	120	1,575	10,274	1,832	6,595	1,602	15,322	15,183	283	2,482	22,236	6,070
CUS	13,090	181	2,248	12,303	2,323	10,135	960	20,867	20,314	59	3,631	33,011	4,598
ASD	8,035	78	1,760	8,119	1,803	6,005	905	13,396	13,251	58	2,280	21,051	3,374
AAL	8,168	146	1,567	8,353	1,818	6,502	904	13,799	13,602	57	2,583	21,070	3,805

Item & Scale Statistics

This section includes updated item descriptive statistics and reliability estimates for each of the Pulse Check Leap-specific scales that make up the Pulse Check survey form. The full text of those items can be found in Appendix C. We limited analysis to responses from students who answered all of the items for the Pulse Check version of that scale. Figure 5 depicts the distribution of mean responses for each of the ten Pulse Check Scales. Additionally, scale reliability information is presented in Table 12. In order to determine if our scales were sufficiently reliable for school and classroom-level aggregate reporting, we chose a cut-off of $\alpha > 0.7$ (Taber, 2018).

HE

The total number of complete responses to the HE Pulse Check Scale was 40,602. Cronbach's α was 0.73, sitting above our a priori threshold of 0.70. Guttman's λ_6 was slightly lower, at 0.68, which is acceptable, especially given the length of the scale. The mean response out of five was 3.31 with a standard deviation of 0.78. Classical item statistics indicate high levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

WCF

A total of 29,009 complete responses were collected for the WCF Pulse Check Scale. The mean response out of five was 3.50 with a standard deviation of 0.84. Cronbach's α was 0.79 and Guttman's λ_6 was 0.71, indicating a high degree of internal consistency. Classical item statistics indicate moderately high levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

RL

A total of 41,938 complete responses were collected for the RL Pulse Check Scale. Cronbach's α was 0.72, which is acceptably above 0.70. Guttman's λ_6 was slightly lower, at 0.64, which is again acceptable, especially given the length of the scale. The mean response out of five was 3.36 with a standard deviation of 0.81. Classical item statistics indicate high levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

REL

We collected 40,435 complete responses to the REL Pulse Check Scale. Cronbach's α was 0.73, which is acceptable. Guttman's λ_6 was slightly lower, at 0.64, which is acceptable, especially given the length of the scale. The mean response out of five was 3.13 with a

standard deviation of 0.93. Classical item statistics indicate moderately high levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

ASO

The total number of complete responses to the ASO Pulse Check Scale was 29,220. Cronbach's α was 0.85 and Guttman's λ_6 was 0.79, which both indicate extremely high degrees of internal consistency in the scale. The mean response out of five was 3.43 with a standard deviation of 0.99. Classical item statistics indicate moderate levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

SCA

A total of 37,744 complete responses were collected for the SCA Pulse Check Scale. Cronbach's α was 0.62 and Guttman's λ_6 was even lower at 0.45, neither of which are acceptable and warrant increasing the scale length. The mean response out of five was 3.37 with a standard deviation of 0.94. Classical item statistics indicate moderate levels of agreement with variability among individual respondents, and item-total correlations are uniform, but slightly lower than other scales.

Due to the lower-than-desired reliability statistics, this scale will be expanded beyond two items in future iterations of the Leaps Pulse Check Scales. Results from piloting indicate that the inclusion of V2_SCA_Q_11 ("How well does your school help students speak out against injustices in society, such as racism?") would strike the best balance of unidimensionality and reliability based upon the EFA and reliability coefficients.

CC

There were 30,788 complete responses to the CC Pulse Check Scale. Cronbach's α was 0.80 and Guttman's λ_6 was 0.73, which both indicate a fairly high degree of internal consistency in the scale. The mean response out of five was 3.40 with a standard deviation of 0.92. Classical item statistics indicate moderate levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

CUS

We collected 41,240 complete responses to the CUS Pulse Check Scale. Cronbach's α was 0.71, which is acceptable. Guttman's λ_6 was slightly lower, at 0.62, which is expected given the length of the scale. The mean response out of five was 3.60 with a standard deviation of

0.78. Classical item statistics indicate moderately high levels of agreement with variability among individual respondents, and item-total correlations are uniformly high.

ASD

The total number of complete responses to the ASD Pulse Check Scale was 26,705. Cronbach's α was 0.70, which is acceptable. Guttman's λ_6 was slightly lower, at 0.65, which is acceptable, especially given the length of the scale. The mean response out of five was 3.21 with a standard deviation of 0.81. Classical item statistics indicate moderately high levels of agreement with variability among individual respondents. The item-total correlations are uniformly high, however, they are slightly lower than some of the scales with stronger internal consistency.

AAL

A total of 27,458 complete responses were collected for the AAL Pulse Check Scale. Cronbach's α was 0.70, which is acceptable. Guttman's λ_6 was slightly lower, at 0.62, which is also acceptable given the length of the scale. The mean response out of five was 3.18 with a standard deviation of 0.77. Classical item statistics indicate moderately high levels of agreement with variability among individual respondents. The item-total correlations for this scale are uniform and moderately high, but slightly lower than for other scales.

Table 11. Final Set of Leaps Pulse Items with Classical Item Statistics

Scale	Item	Mean	SD	Item-Total Correlation
HE	V2_HE_Q_1	3.6	0.97	0.51
	V2_HE_Q_2	3.04	1.1	0.51
	V2_HE_Q_3	3.4	0.99	0.53
	V2_HE_Q_13	3.18	1.13	0.55
WCF	V2_WCF_Q_1	3.23	1.13	0.64
	V2_WCF_Q_2	3.27	1.11	0.65
	V2_WCF_Q_11	3.4	1.09	0.58
RL	V2_RL_Q_1	3.46	1	0.53
	V2_RL_Q_7	3.43	0.98	0.58
	V2_RL_Q_2	3.17	1.06	0.53
REL	V2_REL_Q_1	2.95	1.15	0.52
	V2_REL_Q_3	3.37	1.18	0.55
	V2_REL_Q_2	3.06	1.15	0.58

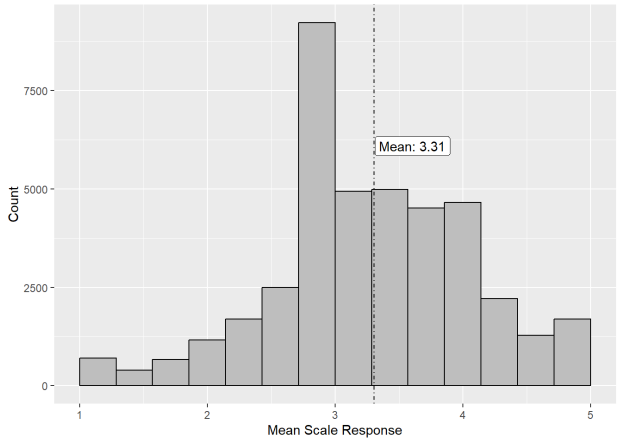
ASO	V2_ASO_Q_2	3.52	1.1	0.7
	V2_ASO_Q_4	3.37	1.12	0.74
	V2_ASO_Q_3	3.4	1.17	0.7
SCA	V2_SCA_Q_16	3.46	1.08	0.45
	V2_SCA_Q_17	3.29	1.12	0.45
CC	V2_CC_Q_1	3.44	1.05	0.65
	V2_CC_Q_4	3.38	1.04	0.64
	V2_CC_Q_2	3.38	1.19	0.63
CUS	V2_CUS_Q_13	3.59	1.01	0.52
	V2_CUS_Q_8	3.65	1	0.52
	V2_CUS_Q_2	3.56	0.92	0.55
ASD	V2_ASD_Q_1	3.66	0.95	0.47
	V2_ASD_Q_3	2.99	1.16	0.49
	V2_ASD_Q_6	3.31	1.11	0.53
	V2_ASD_Q_2	2.9	1.23	0.47
AAL	V2_AAL_Q_1	2.77	1.13	0.49
	V2_AAL_Q_17	3.4	0.97	0.5
	V2_AAL_Q_2	3.09	1.13	0.51
	V2_AAL_Q_15	3.47	1.01	0.45

Table 12. Summary of Item and Reliability Analyses for Finalized Pulse Check Scale

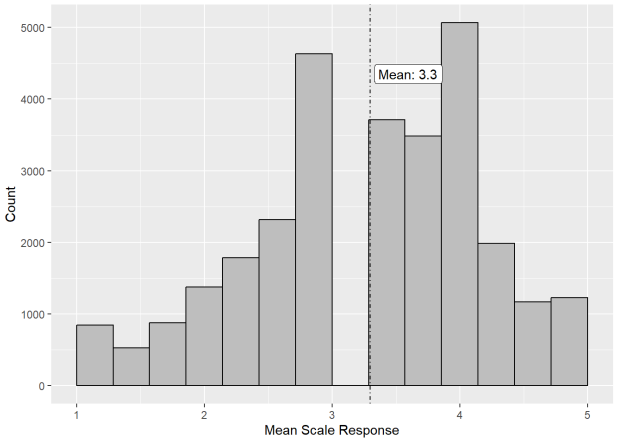
Scale	Cronbach's α	Guttman's λ_6
HE	0.73	0.68
WCF	0.79	0.71
RL	0.72	0.64
REL	0.73	0.64
ASO	0.85	0.79
SCA	0.62	0.45
CC	0.80	0.73
CUS	0.71	0.62
ASD	0.70	0.64
AAL	0.70	0.65

Figure 5. Mean Response Distributions for Each Pulse Subscale

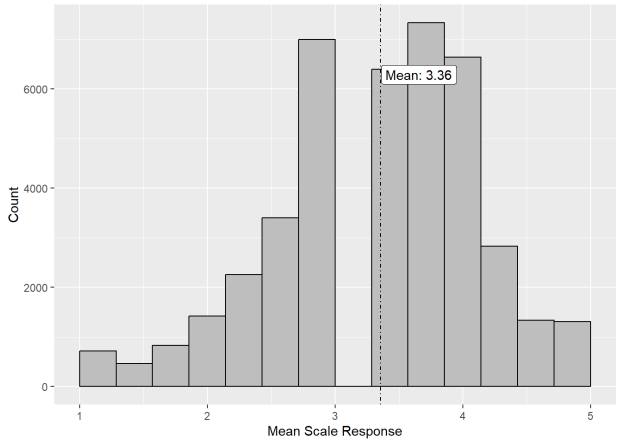
HE WCF



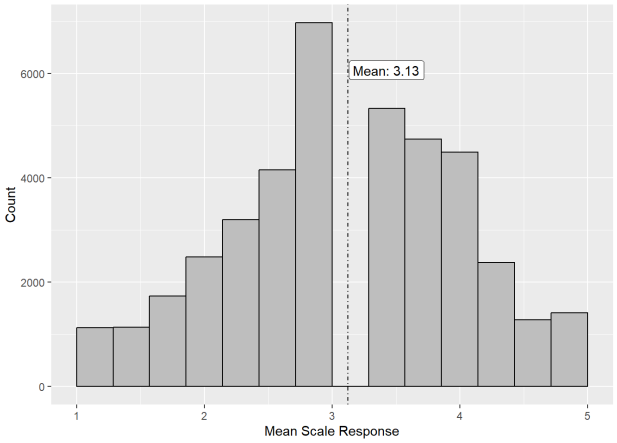
RL



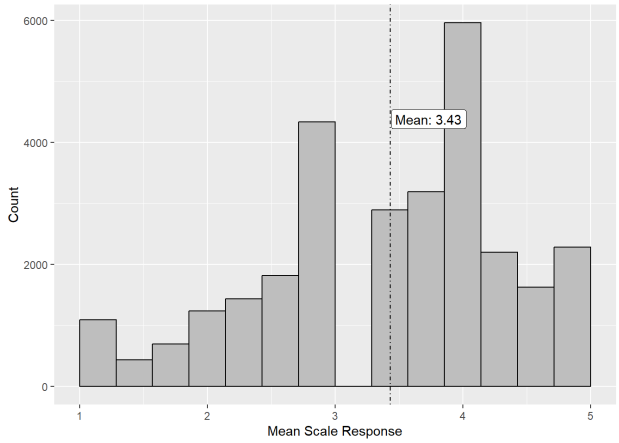
REL



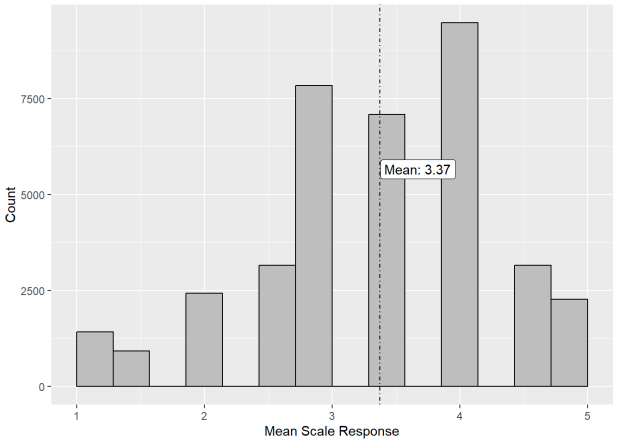
ASO



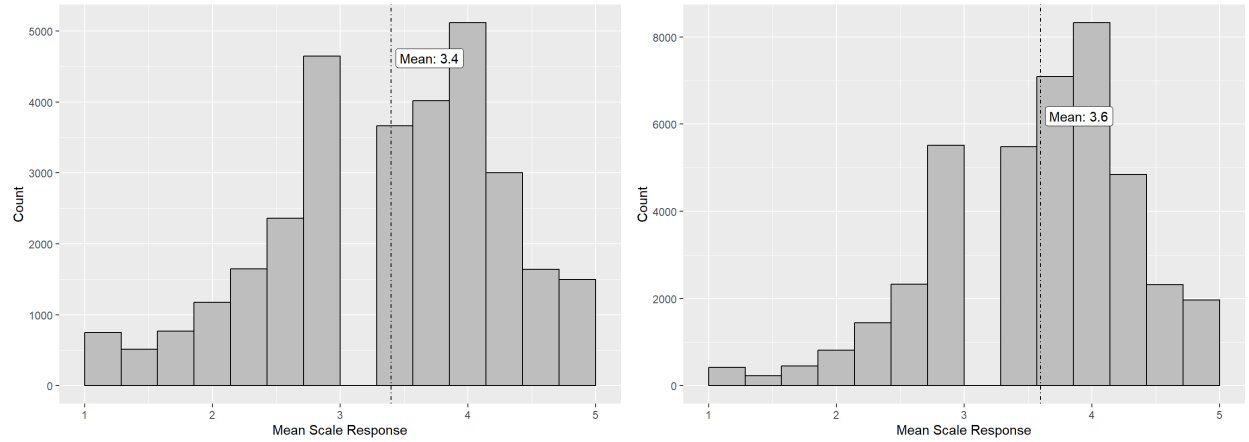
SCA



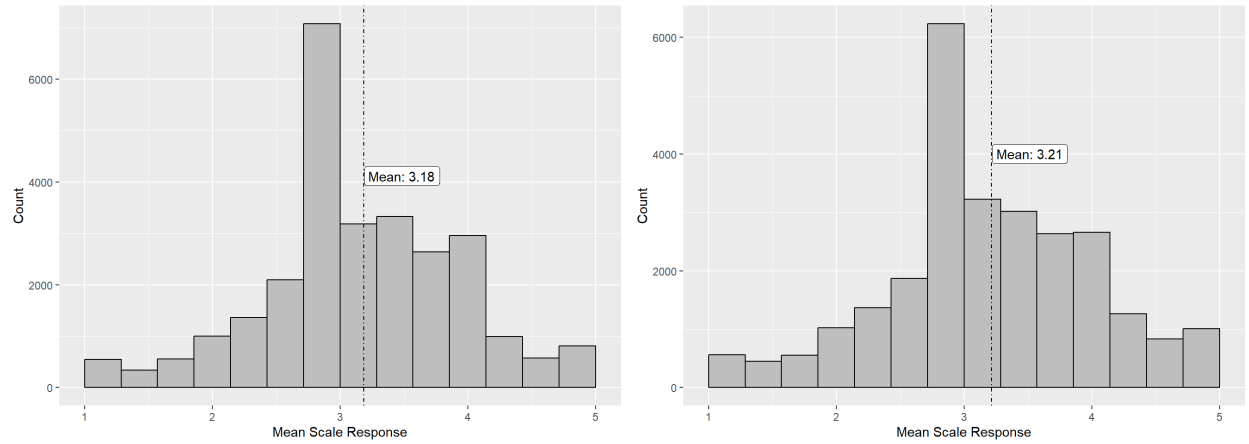
CC



CUS



AAL ASD

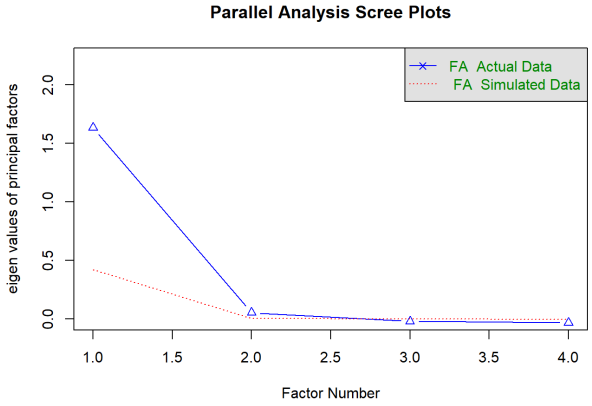


Exploratory Factor Analysis

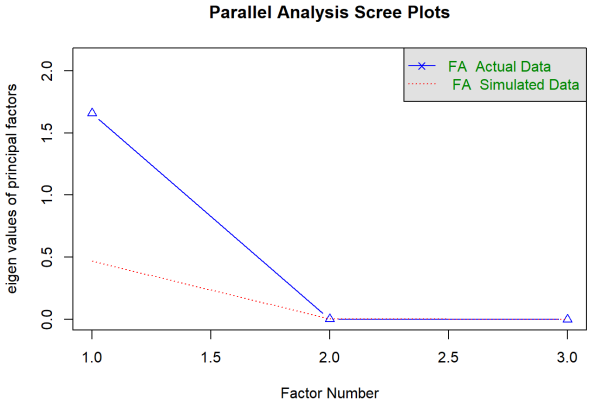
For exploratory factor analysis, we used the same factoring and rotation methods as we did for the Deep Dive analyses. In all cases, visual analysis of the scree plots indicated that a greater-than-one factor structure was desirable (Figure 6). Thus, after parallel analysis, our next step was to separately fit unidimensional and two-dimensional factor models for each subscale. However, in all cases, the short length of the subscale led to negative degrees of freedom for a solution with two or more factors. As a result, we focused mainly on the fit of the unidimensional model. We used principal axis factoring and polychoric correlations for the analyses, similar to the above analysis for the full Deep Dive Scales. As can be seen in Table 13, for each subscale of the Leaps Pulse Check Scales, RMSR fell below 0.08, indicating that the unidimensional solution had a strong model fit (Hornsby et al., 2021).

Figure 6. Scree Plots for Each Pulse Subscale

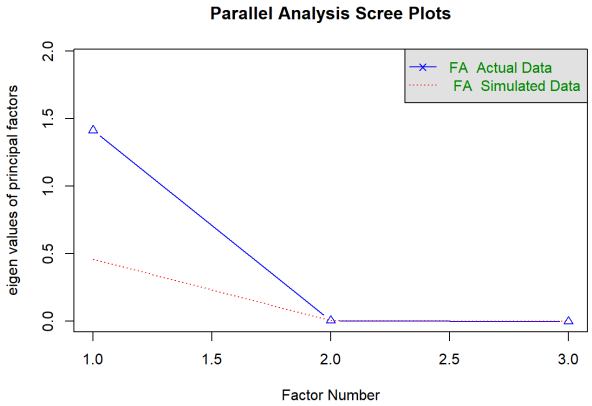
HE WCF



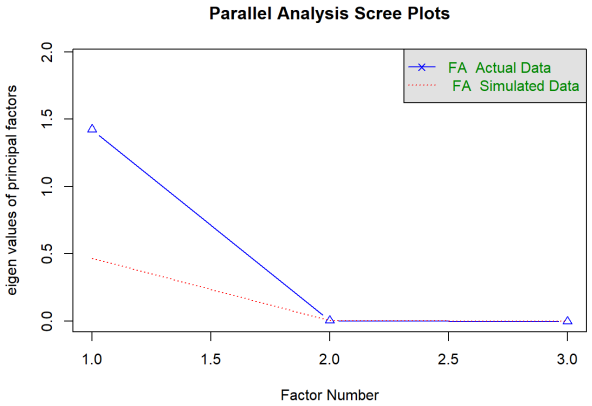
RL



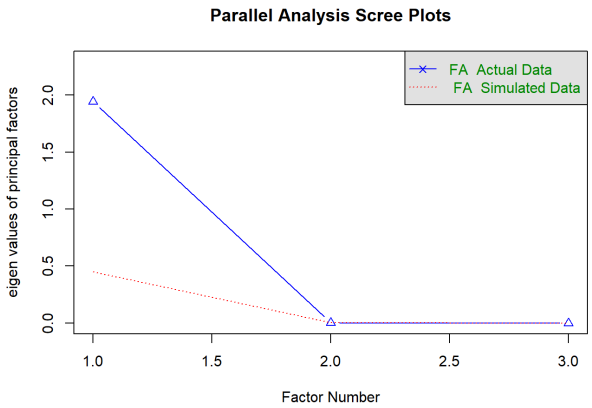
REL



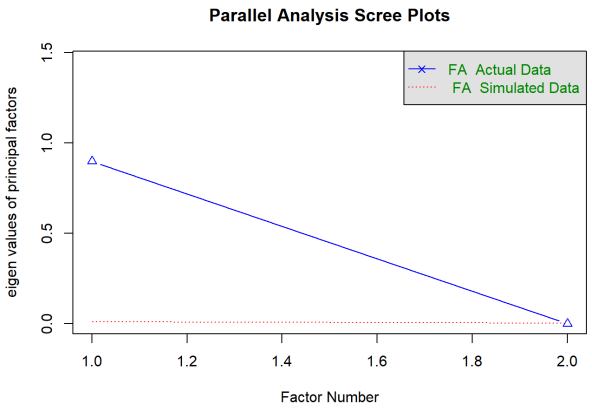
ASO



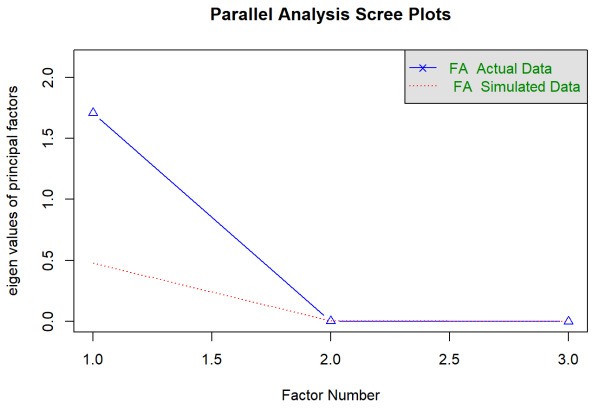
SCA



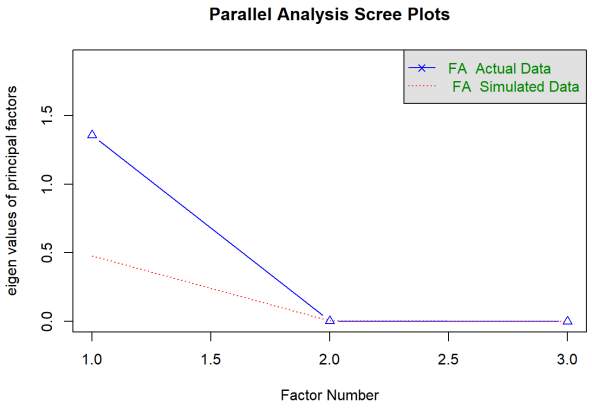
CC



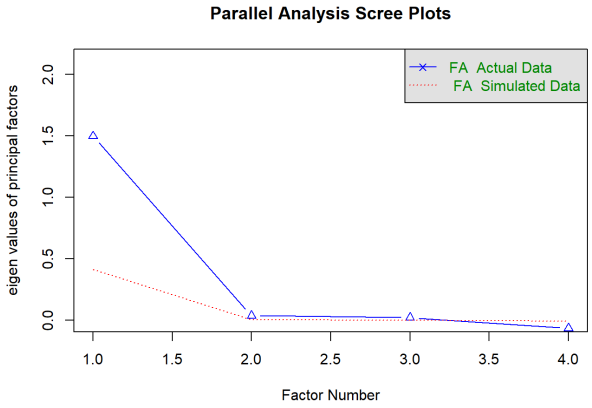
CUS



AAL



ASD



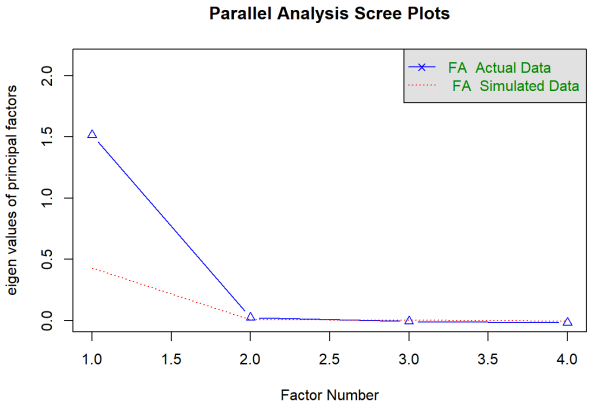


Table 13. Unidimensional RMSR Values for Each Pulse Check

Scale	RMSR
HE	0.019
WCF	0.001
RL	0.002
REL	0.001
ASO	0.001
SCA	0.000
CC	0.000
CUS	0.000
AAL	0.023
ASD	0.009

Relationships with Other Variables

The Leaps Pulse Check Scale administrations also include two items intended to capture students' attitudes toward school in general. These items and identifiers were:

- V2_GEN_1: Overall, most of the time, I love my school.
- V2_GEN_2: Overall, most of the time, I'm learning a lot in my school.

We expected that every subscale would be fairly strongly associated with both of these variables, but that Leaps related to learning directly would relate more to V2_GEN_2, while more affective Leaps would relate more to V2_GEN_1. To test this, we calculated a mean score for each subscale (based on the subscale recommendations indicated above), and then calculated the correlation between that score and each of the "general" Leap items. Table 14 summarizes these findings.

Table 14. Correlations of Pulse Check Subscales with Love of School and Learning in School

Scale	Correlation with "Love School"	Correlation with "Learning in School"
HE	.53	.59
WCF	.55	.55
RL	.47	.56
REL	.54	.58
ASO	.52	.48
SCA	.34	.40
CC	.53	.48
CUS	.45	.58
ASD	.47	.53
AAL	.47	.52

Correlations tended to be similar across both general items, but we do see some evidence of stronger associations with V2_GEN_2, the general item dealing with learning, for CUS and RL, both of which do pertain to learning. Similarly, the scales less related to learning, such as WCF, were more strongly associated with the general item asking about loving school.

Spanish-Language Leaps Pulse Check

A Spanish version of the Pulse Check was made available for Fall 2022, and students were able to opt into this version of the scale using a dropdown menu. We analyzed the item

response data from Fall 2022 to understand the extent to which the Spanish version of the Pulse Check was producing scores comparable to those from the previously-investigated English-language version.

Ideally, it should be possible to interpret responses to the Spanish-language version of the Pulse Check identically to how responses to the English-language version are interpreted. Essentially, it should not matter at all which language a student chose when responding to the Pulse Check. To test this assumption, we investigated the factor structure of the Pulse Check Scale to see if the structural equation model (confirmatory factor model) holds for the Spanish-language version as well.

We began our investigation by fitting a multigroup CFA, one factor per Leap, with correlated factors. The multigroup CFA immediately indicated problems with the proposed factor structure for the Spanish-language version of the test in the form of a non-positive definition factor covariance matrix. This made interpretation of the multigroup model problematic, so we turned to fitting the model separately by language. These separate CFAs indicated that the specified factor solution fit the English-language responses quite well, but was not a viable solution for the Spanish-language version (noting that the factor covariance matrix was not positive definite, fit was also poor.)

At this point, it was apparent that the same factor structure was not appropriate for the Spanish-language Pulse Check as for the English version. The next step was to conduct an EFA for the Spanish-language version to try to understand if an alternate, but interpretable, factor structure was emerging. At this point, we also investigated the reliability of each Leap-specific Pulse Check subscale. As summarized in Table 16, reliability was lower than acceptable, with α falling below 0.70 in every case.

Table 15. Summary of Fall 2022 Analytic Sample for Spanish Invariance Analyses
Scale Analysis

Grade	Spanish language N	English language N
All	237	6835
3	30	292
4	31	380
5	28	362
6	19	1289
7	37	1623

8	31	1652
9	24	427
10	22	401
11	11	233
12	4	176

Table 16. Reliability estimates for Spanish-language Pulse Check subscales

Leap	Cronbach's α
HE	0.66
WCF	0.64
RL	0.60
REL	0.51
ASD	0.57
SCA	0.69
CC	0.68
CUS	0.62
ASO	0.54
AAL	0.46

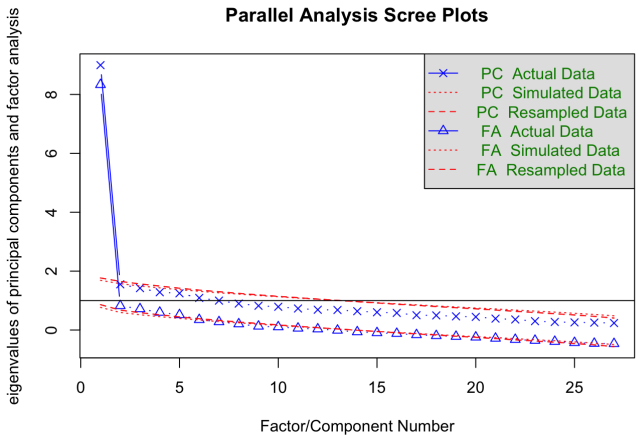


Figure 7. Parallel Analysis of Spanish-Language Pulse Check

Parallel analysis indicated that there was evidence of five factors, as shown in Figure 7. We fit solutions with up to five factors, using the same estimation and rotation methods as the English-language EFAs of the Pulse Check. No version with more than one factor produced

an interpretable factor structure. Nearly every scale had at least one item load on multiple factors, or had items loading on at least two distinct factors, as shown below in the five-factor solution in Table 17.

Table 17. Five-factor EFA of Spanish-language Pulse Check

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
V2_HE_Q_1		0.32			
V2_HE_Q_2		0.55			
V2_HE_Q_3	0.5				
V2_HE_Q_13	0.42		0.29		
V2_WCF_Q_1	0.49	0.28			
V2_WCF_Q_2		0.41			
V2_RL_Q_1	0.32				
V2_RL_Q_2	0.31				0.36
V2_RL_Q_7	0.29				
V2_REL_Q_1					0.52
V2_REL_Q_2	0.34				0.5
V2_REL_Q_3			0.3		
V2_ASO_Q_2	0.77				
V2_ASO_Q_3	0.6				
V2_SCA_Q_16		0.75			
V2_SCA_Q_17		0.9	0.26		
V2_CC_Q_1	0.26			0.59	
V2_CC_Q_2				0.63	
V2_CUS_Q_2			0.27	0.44	
V2_CUS_Q_8			0.6		
V2_CUS_Q_13			0.59		
V2_ASD_Q_1			0.66		
V2_ASD_Q_2				0.41	
V2_ASD_Q_3			0.46	0.33	
V2_AAL_Q_1		0.31			0.41
V2_AAL_Q_2		0.28			
V2_AAL_Q_15					0.28

The inconsistent factor structure and low reliability of each subscale indicate that a revision to the Spanish version of the Pulse Check is warranted. However, moving beyond classical

test statistics, we acknowledge that it is necessary to provide a Spanish-language version of the Pulse Scales. As is shown in both Table 5 and 10, Latino students make up a large portion of the students served by the Leaps.

Subsequent to receiving this analysis, Transcend worked with a professional translation service to update the Spanish-language questions and is now testing them in the field. We will share new analyses of these updated items when sufficient data has been collected to enable the full suite of desired psychometric tests.

Applied Case Study

Case studies can offer real-world evidence to support the effectiveness of the Leaps Student Voice Survey. To date, one case study has been conducted with the Leaps. This case study took place in the Salem (MA) Public Schools between 2022 and 2023. Students were given either Leaps Deep Dives or Pulse Checks, then educators changed the learning environment in response to the data, and then administered the same Leaps survey to measure changes in student perspective. Data on the Leaps survey increased over the course of the year, as did other data such as attendance and parent satisfaction. This case study supports the theory of action for the Leaps Student Voice Survey.

In the wake of the 2020 COVID pandemic, Salem Public Schools sought to reinvent the middle-school experience. Students were reporting low engagement in school and an increase in mental health issues. Thus, Salem PS partnered with WPS Institute, a nonprofit dedicated to educational innovation, to take community and parent feedback and design a new student-centered middle school experience. The newly developed pilot featured hands-on learning supported by partnerships with community organizations. To understand how students were experiencing the pilot, Salem PS and WPS partnered with Transcend to administer the Leaps Pulse Check Scales on multiple occasions throughout the first year of the pilot.

Three high-level goals, Empowered, Connected, and Unbounded Learning, were actualized through Personalized, Community-based, and Hands-on Learning. Each of these goals aligned closely with specific Leaps:

- Personalization allows students to set the speed of their learning, choose their own learning goals, and reflect on their progress with chosen mentors. Progress in this area was measured through the ASD and REL Leaps.

- Community-based learning allows students to immerse themselves in the Salem community, creating a tangible link between what they are learning in school and the world around them. Progress in this area was measured with the CC and WCF Leaps.
- Hands-on learning allows students to grow through design studios, where they are briefly introduced to a topic and then design, sketch, plan, and refine a final project. Progress in this area was measured with the AAL Leap.

A total of forty-five grade 8 middle-school students were selected from a lottery to participate in the pilot. The pilot was mostly representative of the district demographically, although students qualifying for special education services were oversampled in the pilot population. The sample was largely white (47%), followed by Black (23%) and Latino (12%) students. The majority of participants (67%) receive free or reduced-priced lunch. Although the Salem PS pilot focused on the five aforementioned Leaps, the entirety of the Pulse Check Scale was administered to participants at the beginning and end of the pilot, roughly eight months apart.

The results of these administrations are depicted in Table 18 and Figure 8. For each of the targeted scales (WCF, REL, CC, ASD, AAL), paired t-tests demonstrated a statistically significant increase in total score, providing evidence of convergent validity to these scales. Convergent validity refers to the degree to which different methods or measures intended to assess the same construct are correlated or agree with each other. On the other hand, of the non-targeted scales, only HE showed statistically significant gains, lending evidence of discriminant validity to the RL, ASO, SCA, and CUS Pulse Check Scales, which the intervention was not intended to target.

Table 18. Summary of Pilot Studies

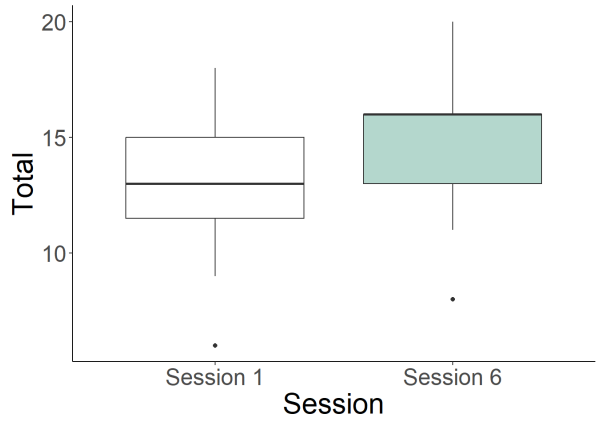
Pulse Scale	Session	N	Mean	SD	p
HE	First	39	13.15	2.63	
	Last	39	14.98	2.59	.002**
WCF	First	38	9.59	2.63	
	Last	41	11.12	2.24	.008**
RL	First	39	10.20	2.10	
	Last	41	11.07	2.19	.06
REL	First	39	8.36	2.82	
	Last	37	10.31	2.40	.001***
ASO	First	39	10.42	2.74	

	Last	41	11.07	2.52	.61
SCA	First	39	7.11	1.64	
	Last	41	7.60	1.35	.19
CC	First	38	9.76	2.73	
	Last	41	11.37	1.94	.013*
CUS	First	38	10.61	2.02	
	Last	40	11.31	2.34	.12
ASD	First	39	12.60	2.75	
	Last	41	14.14	2.53	.015*
AAL	First	39	12.75	3.19	
	Last	41	14.81	2.91	.005**

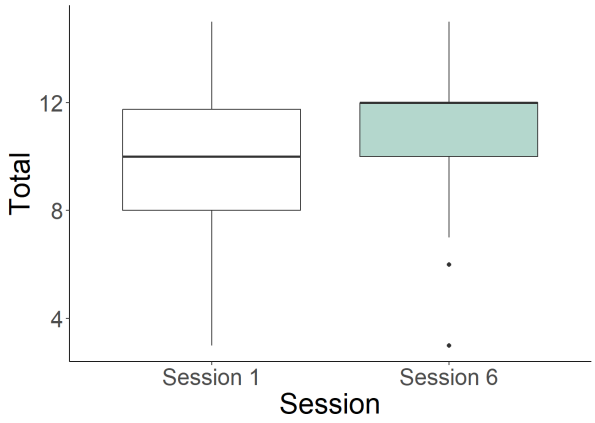
Note. * indicates significance at the p < 0.05 level, ** indicates significance at the p < 0.01 level, and *** at the p < .001 level

Figure 8. Salem PS Leaps Scores Before and After Intervention

HE

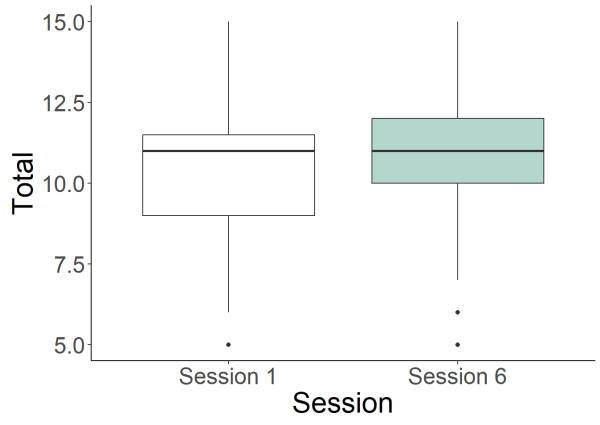


WCF

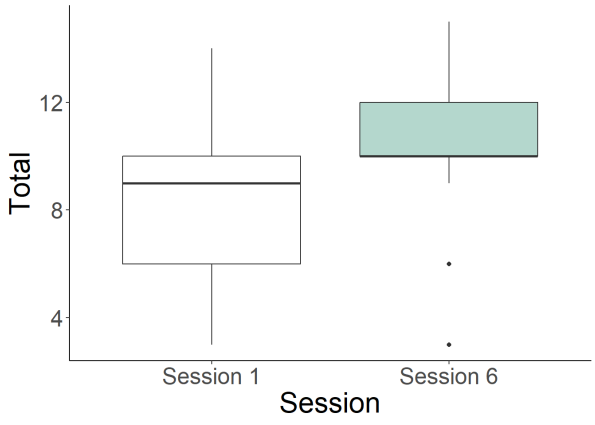


RL

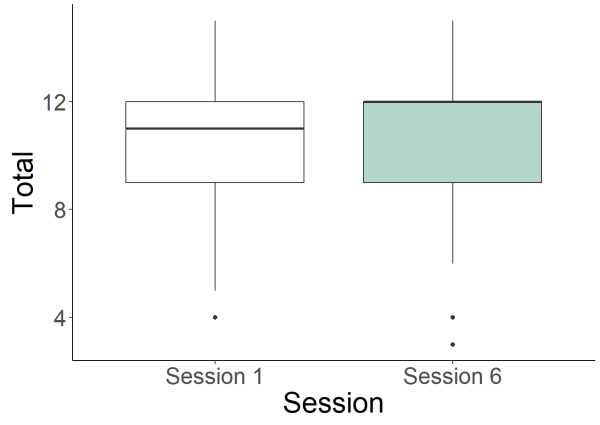
REL



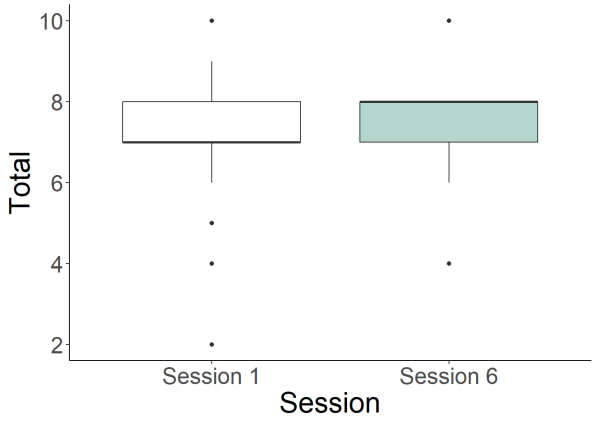
ASO



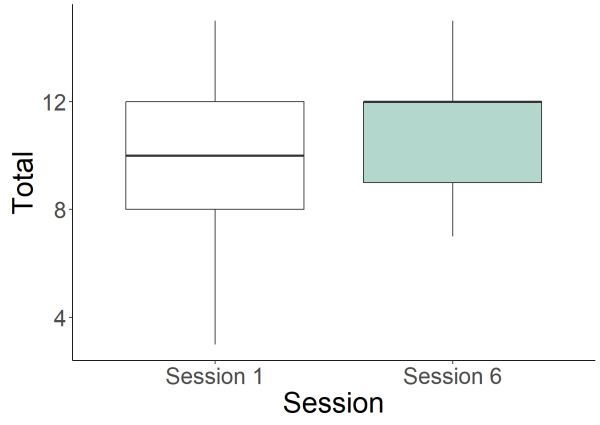
SCA



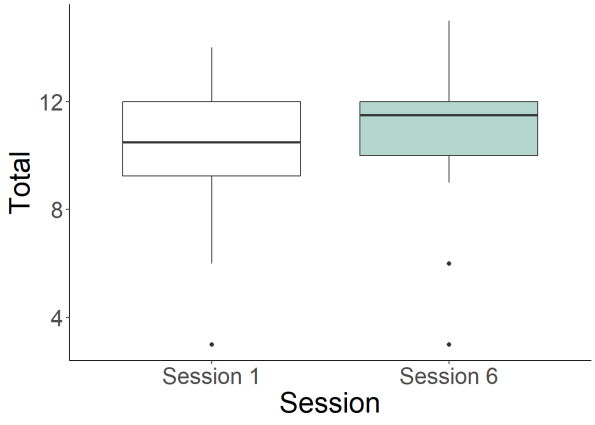
CC



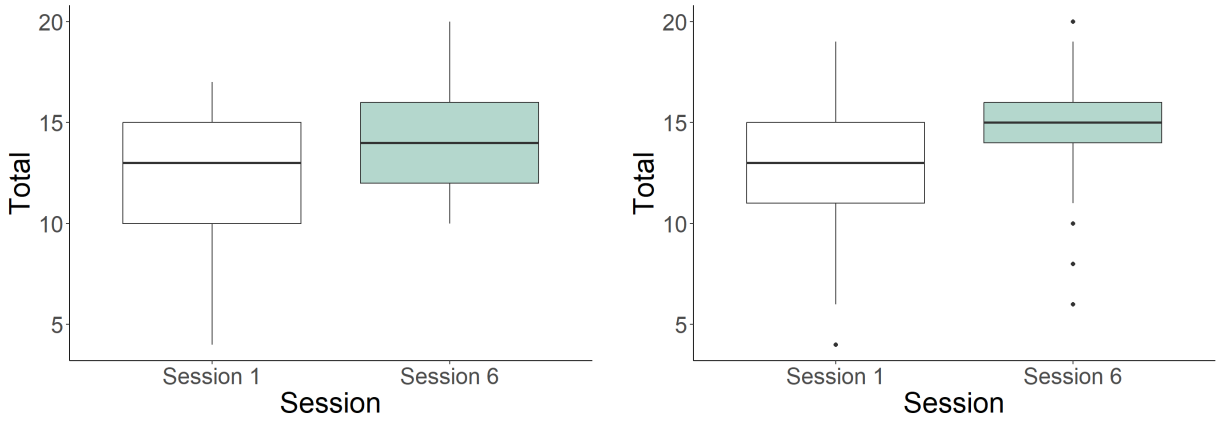
CUS



ASD



AAL



The Salem PS pilot aligns closely with certain Leaps scales (WCF, REL, CC, ASD, AAL) and less so with others (HE, RL, ASO, SCA, CUS). In practice, after participating in the pilot MS for a year, students saw increases in the averages of the target scales and little-to-no change in the non-targeted scales. Consequently, empirical convergent and discriminant evidence supports the claims that the WCF, REL, CC, ASD, and AAL Leaps Pulse Check Scales are all sensitive to changes in the constructs they are intended to measure and the RL, ASO, SCA, and CUS scales do not measure what they are not intended to measure. This provides strong validity evidence in support of the intended purposes and uses of the Leaps Student Voice Survey including repeatedly administering specific Pulse Check Scales aligned with project goals to understand if a pilot is progressing.

References

Aldridge, J. M., & Fraser, B. J. (2008). Technology-Rich Outcomes-Focused Learning Environment Inventory (TROFLEI). In *Outcomes-Focused Learning Environments* (pp. 95-100). Brill Sense.

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, *44*(5), 427-445.

Assor, A., Kaplan, H., & Roth, G. (2002). Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviours predicting students' engagement in schoolwork. *British journal of educational psychology*, *72*(2), 261-278.

Burns, M. K., Vance, D., Szadokierski, I., & Stockwell, C. (2006). Student needs survey: A psychometrically sound measure of the five basic needs. *International Journal of Reality Therapy*, *25*(2), 4.

Calderón Garrido, C., Navarro González, D., Lorenzo Seva, U., & Ferrando Piera, P. J. (2019). Multidimensional or essentially unidimensional? A multi-faceted factoranalytic approach for assessing the dimensionality of tests and items. *Psicothema, 31*(4), 450-457.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245-276.

Chai, C. S., Deng, F., Tsai, P. S., Koh, J. H. L., & Tsai, C. C. (2015). Assessing multidimensional students' perceptions of twenty-first-century learning practices. *Asia Pacific Education Review*, *16*(3), 389-398.

Ferguson, R. (2010, October 14). Student perceptions of teaching effectiveness. Discussion brief from the National Center for Teacher Effectiveness and the Achievement Gap Initiative, Harvard University, Cambridge, MA.

Frymier, A. B., & Shulman, G. M. (1995). "What's in it for me?": Increasing content relevance to enhance students' motivation. *Communication Education*, *44*(1), 40-50.

Hart, H., Young, C., Chen, A., Kheraj, N., & Allensworth, E. M. (2021). " 5Essentials" Survey in CPS: School Improvement and School Climate in High Poverty Schools. Research Report. *University of Chicago Consortium on School Research*.

Hornsby, B. W., Camarata, S., Cho, S. J., Davis, H., McGarrigle, R., & Bess, F. H. (2021). Development and validation of the Vanderbilt Fatigue Scale for Adults (VFS-A). *Psychological Assessment, 33*(8), 777.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*(1), 141-151.

Massachusetts Consortium for Innovative Education Assessment. (2021). School Quality Measures: Student and Teacher Survey Scales. Retrieved from: https://www.mciea.org/school-quality-measures.html

MCIEA see Massachusetts Consortium for Innovative Education Assessment

Panorama. (2021). Panorama Social-Emotional Learning Survey. Retrived from: https://www.panoramaed.com/social-emotional-learning-sel

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*, 1273-1296.

United States Department of Education. (2017). ED School Climate Surveys (EDSCLS). Retrieved from: https://safesupportivelearning.ed.gov/edscls/administration

Young, M. R., Caudill, E. M., & Murphy, J. W. (2008). Evaluating experiential learning activities. *Journal for Advancement of Marketing Education*, *13*(Winter), 28-40.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores.* Directorate of Human Resources Research and Evaluation, Department of National Defense.

Appendices

Appendix A. Full Text for Leaps Deep Dive Scale Items

Scale	Item	Item Text
HE	V2_HE_Q_1	At my school, it feels like I'm expected -- and supported -- to learn a ton.
	V2_HE_Q_2	At my school, people don't give up when the work gets hard.
	V2_HE_Q_3	I feel like I have access to all of the opportunities my school offers.
	V2_HE_Q_6	Adults at my school work hard to make sure that all students are learning.
	V2_HE_Q_7	Adults at my school pay attention to all students, not just the top students.
	V2_HE_Q_8	At my school, my questions get just as much attention as other students' questions.
	V2_HE_Q_9	I am treated the same as other kids at my school.
	V2_HE_Q_10	I receive the same encouragement from adults at my school as other kids do.
	V2_HE_Q_11	I get the same opportunity to contribute to class discussions as others.
	V2_HE_Q_12	I get the same opportunity to answer questions as other kids at my school.
	V2_HE_Q_13	When you feel like giving up on a difficult task, how likely is it that someone at your school will help you keep trying?
WCF	V2_WCF_Q_1	At my school I learn to understand my emotions.
	V2_WCF_Q_2	At my school I learn how to figure out who I am as a person.

	V2_WCF_Q_3	At my school, I'm growing in many different ways -- such as my physical health, my emotional well-being, and understanding myself better.
	V2_WCF_Q_4	At my school I learn how to keep my body healthy.
	V2_WCF_Q_5	At my school I learn how to eat well.
	V2_WCF_Q_6	At my school I learn how to stay physically active.
	V2_WCF_Q_7	At my school I learn how to cope with stress in my life.
	V2_WCF_Q_8	At my school I learn how to understand the perspectives of others.
	V2_WCF_Q_9	At my school I learn how to recognize my strengths.
	V2_WCF_Q_10	At my school we talk about feelings and emotions.
	V2_WCF_Q_11	At my school, everyone wants me to not just learn, but also be happy and healthy in mind and body.
RL	V2_RL_Q_1	At my school we use our thinking skills, rather than just memorizing things.
	V2_RL_Q_2	At my school I get to develop my own ideas.
	V2_RL_Q_3	At my school I think hard, analyze ideas, and solve complicated problems.
	V2_RL_Q_4	At my school I explain my answers -- why I think what I think.
	V2_RL_Q_5	At my school we learn a lot almost every day.
	V2_RL_Q_6	At my school we learn to correct our mistakes.
	V2_RL_Q_7	At my school we have time to explain our ideas.
	V2_RL_Q_8	At my school I think about other possible ways of understanding what I am learning.

	V2_RL_Q_9	At my school I consider different opinions to see which one makes more sense.
	V2_RL_Q_10	At my school I get to come up with new ideas.
	V2_RL_Q_11	At my school I get to be creative.
REL	V2_REL_Q_1	At my school what we learn is often connected to life outside the classroom.
	V2_REL_Q_2	At my school I get to learn things I'm interested in.
	V2_REL_Q_3	At my school what I'm learning matters a lot to me.
	V2_REL_Q_4	Adults at my school help me see how what I am learning relates to my life.
	V2_REL_Q_5	Adults at my school care whether what I am learning is interesting to me.
	V2_REL_Q_6	Adults at my school talk about the connections between what we study and what happens in real life.
	V2_REL_Q_7	Adults at my school help me relate my personal experiences to what I am learning.
	V2_REL_Q_8	Adults at my school help me connect the content with things I learned in the past.
	V2_REL_Q_9	At my school I can apply what I'm learning to my own interests.
	V2_REL_Q_10	At my school we talk about current events related to what we are learning.
	V2_REL_Q_11	At my school I apply the knowledge I have to solve real-life problems.
ASO	V2_ASO_Q_1	At your school how often do you see people like you represented in what you study?
	V2_ASO_Q_2	At my school I feel proud of who I am.

	V2_ASO_Q_3	I can be myself at my school.
	V2_ASO_Q_4	At my school it feels like being yourself is a great thing. I feel safe and appreciated for who I am.
	V2_ASO_Q_5	At your school how often do you see many different kinds of people represented in what you study?
	V2_ASO_Q_6	My school provides instructional materials (e.g., textbooks, handouts) that reflect my cultural background, ethnicity, and identity.
	V2_ASO_Q_7	How valued do you think all students' home cultures and languages are in the school curriculum?
	V2_ASO_Q_8	How valued do you think your home culture and language are in the school curriculum?
	V2_ASO_Q_10	People at my school appreciate me for who I am.
	V2_ASO_Q_11	At my school I learn new things about my culture.
SCA	V2_SCA_Q_2	How well does your school help students speak out against injustices in society, such as racism?
	V2_SCA_Q_3	How often at your school do you learn about people from different races, ethnicities, or cultures?
	V2_SCA_Q_4	How often at your school do you think about what someone of a different race, ethnicity, or culture experiences?
	V2_SCA_Q_5	At your school, how often are you encouraged to think more deeply about injustices and inequities in society with other students?
	V2_SCA_Q_6	How often do students at your school have important conversations about injustices and inequities in society, even when they might be uncomfortable?
	V2_SCA_Q_7	When there are major events related to race, how often do adults at your school talk about them with students?

	V2_SCA_Q_8	How confident are you that students at your school can have honest conversations with each other about injustices and inequities in society, such as racism?
	V2_SCA_Q_9	How comfortable are you sharing your thoughts about race-related topics with other students at your school?
	V2_SCA_Q_13	How often do students at your school have important conversations about race, even when they might be uncomfortable?
	V2_SCA_Q_16	At my school we learn about racism in the United States.
	V2_SCA_Q_17	At my school we take action to fix problems in society, such as racism and discrimination.
CC	V2_CC_Q_1	I feel part of the community at my school. There are a lot of people who know and care about me.
	V2_CC_Q_2	Overall, how much do you feel like you belong at your school?
	V2_CC_Q_3	At my school adults are there for me when I need them.
	V2_CC_Q_4	At my school I feel included by other students.
	V2_CC_Q_5	At my school students like me the way I am.
	V2_CC_Q_6	At my school I feel comfortable.
	V2_CC_Q_7	At my school I work together with other people to learn new things.
	V2_CC_Q_8	At my school I get helpful comments about my work from other students.
	V2_CC_Q_9	How well do people at your school understand you as a person?
	V2_CC_Q_10	How connected do you feel to the people at your school?
CUS	V2_CUS_Q_1	At my school I get to work at my own speed.

	V2_CUS_Q_2	At my school I do work that meets me where I am in my learning.
	V2_CUS_Q_3	My school works just right for me. I am able to learn at my own pace and in my own way
	V2_CUS_Q_4	At my school I get specific suggestions about how I can improve my work.
	V2_CUS_Q_5	At my school I get extra help if I need it.
	V2_CUS_Q_6	At my school students who work faster, move on to the next topic
	V2_CUS_Q_7	At my school when I am doing something that interests me, I have enough time to finish it.
	V2_CUS_Q_8	At my school I have the resources I need to support my learning.
	V2_CUS_Q_9	Adults at my school notice if I have trouble learning something.
	V2_CUS_Q_10	Adults at my school explain things in a different way if I don't understand something.
	V2_CUS_Q_11	Adults at my school give me individual attention when I need it.
	V2_CUS_Q_12	Adults at my school understand how I learn best.
	V2_CUS_Q_13	At my school I am able to catch up if I am behind.
	V2_CUS_Q_14	The learning materials used by my school meet my needs.
	V2_CUS_Q_15	The work I do at my school allows me to really show what I know.
ASD	V2_ASD_Q_1	At my school I have goals for my learning, and I have choices about how I pursue those goals.

	V2_ASD_Q_2	At my school I feel like I have a say about what happens to me.
	V2_ASD_Q_3	At my school I can choose how to do my work.
	V2_ASD_Q_4	At my school I have choices on different ways to complete assignments.
	V2_ASD_Q_5	At my school I can choose to study topics that interest me.
	V2_ASD_Q_6	Adults at my school respect my ideas and suggestions.
	V2_ASD_Q_7	Adults at my school ask us if there are things we would like to change in the way we study.
	V2_ASD_Q_8	Adults at my school encourage me to work in my own way.
	V2_ASD_Q_9	Adults at my school are open to suggestions from students.
AAL	V2_AAL_Q_1	At my school we spend time learning outside of our school building.
	V2_AAL_Q_2	I interact with people outside my school to help me learn.
	V2_AAL_Q_3	My school organizes field trips.
	V2_AAL_Q_4	My school offers community service opportunities.
	V2_AAL_Q_5	My school values the activities I do outside of school.
	V2_AAL_Q_6	At my school I have opportunities to learn from lots of people, not just the teachers.
	V2_AAL_Q_8	My school hosts events before school or after the school day ends.
	V2_AAL_Q_9	We spend time in the community outside of my school.
	V2_AAL_Q_10	At my school, I have opportunities to learn from local community members.

	V2_AAL_Q_11	Adults at my school are interested in what I am learning outside of school.
	V2_AAL_Q_12	Adults at my school are interested in my experiences at home.
	V2_AAL_Q_13	I am able to access my school online.
	V2_AAL_Q_14	If my school was closed for a day, I would still be able to make progress on my schoolwork.
	V2_AAL_Q_15	I can make progress on my learning when I am not at my school.
	V2_AAL_Q_16	I have a way to communicate with my teachers when I am not at my school.
	V2_AAL_Q_17	My school encourages and gives me the tools to be learning everywhere, from everyone, all the time.

Note. References to “my school”, “your school”, or “school” could be replaced with actual organization names in practice.

Appendix B: DIF Statistics for All Leaps Deep Dive Scales

In tables below, * indicates significance at the $p < 0.05$ level

HE scale

DIF by Gender

Item	Pseudo-r^2 difference	DIF effect size
V2_HE_Q_1	0.004	trivial
V2_HE_Q_10	0	trivial
V2_HE_Q_11	-0.001	trivial
V2_HE_Q_12	-0.005	trivial
V2_HE_Q_13	0.002	trivial
V2_HE_Q_2	0.003*	trivial
V2_HE_Q_3	0.003	trivial
V2_HE_Q_6	0	trivial
V2_HE_Q_7	0.008*	trivial
V2_HE_Q_8	0	trivial
V2_HE_Q_9	-0.004	trivial

DIF by Race: Latino/Hispanic vs. Black

Item	Pseudo-r^2 difference	DIF effect size
V2_HE_Q_1	0.001	trivial
V2_HE_Q_10	0.002	trivial
V2_HE_Q_11	0.007*	trivial
V2_HE_Q_12	0.002	trivial
V2_HE_Q_13	-0.003	trivial
V2_HE_Q_2	0	trivial
V2_HE_Q_3	0.002	trivial
V2_HE_Q_6	0.002	trivial
V2_HE_Q_7	0.004	trivial
V2_HE_Q_8	-0.001*	trivial
V2_HE_Q_9	0.003*	trivial

DIF by Race: Latino/Hispanic vs. White

Item	Pseudo-r^2 difference	DIF effect size
V2_HE_Q_1	-0.004	trivial

V2_HE_Q_10	-0.009*	trivial
V2_HE_Q_11	0.008*	trivial
V2_HE_Q_12	0.002	trivial
V2_HE_Q_13	-0.012*	trivial
V2_HE_Q_2	-0.005*	trivial
V2_HE_Q_3	0	trivial
V2_HE_Q_6	-0.008	trivial
V2_HE_Q_7	0	trivial
V2_HE_Q_8	0.001*	trivial
V2_HE_Q_9	-0.001*	trivial

DIF by Race: Latino/Hispanic vs. Asian

Item	Pseudo-r^2 difference	DIF effect size
V2_HE_Q_1	-0.002	trivial
V2_HE_Q_10	-0.002	trivial
V2_HE_Q_11	0.004*	trivial
V2_HE_Q_12	0.001	trivial
V2_HE_Q_13	0*	trivial
V2_HE_Q_2	0.002	trivial
V2_HE_Q_3	-0.002	trivial
V2_HE_Q_6	-0.003	trivial
V2_HE_Q_7	-0.001	trivial
V2_HE_Q_8	0*	trivial
V2_HE_Q_9	-0.011*	trivial

WCF Scale

DIF by Gender

Item	Pseudo-r^2 difference	DIF effect size
V2_WCF_Q_1	0	trivial
V2_WCF_Q_10	0.004*	trivial
V2_WCF_Q_11	0.001*	trivial
V2_WCF_Q_2	0.001	trivial
V2_WCF_Q_3	0	trivial
V2_WCF_Q_4	-0.004	trivial
V2_WCF_Q_5	-0.001	trivial

V2_WCF_Q_6	-0.001	trivial
V2_WCF_Q_7	0.001*	trivial
V2_WCF_Q_8	0.004*	trivial
V2_WCF_Q_9	-0.003	trivial

DIF by Race: Latino/Hispanic vs. Black

Item	Pseudo-r^2 difference	DIF effect size
V2_WCF_Q_1	0.002	trivial
V2_WCF_Q_10	0.002	trivial
V2_WCF_Q_11	0.004	trivial
V2_WCF_Q_2	0	trivial
V2_WCF_Q_3	0	trivial
V2_WCF_Q_4	-0.008*	trivial
V2_WCF_Q_5	0.007*	trivial
V2_WCF_Q_6	0.007*	trivial
V2_WCF_Q_7	-0.004	trivial
V2_WCF_Q_8	0.003	trivial
V2_WCF_Q_9	0.001*	trivial

DIF by Race: Latino/Hispanic vs. White

Item	Pseudo-r^2 difference	DIF effect size
V2_WCF_Q_1	0.006*	trivial
V2_WCF_Q_10	-0.002	trivial
V2_WCF_Q_11	0.001	trivial
V2_WCF_Q_2	0	trivial
V2_WCF_Q_3	0.001	trivial
V2_WCF_Q_4	-0.001	trivial
V2_WCF_Q_5	0	trivial
V2_WCF_Q_6	-0.001	trivial
V2_WCF_Q_7	-0.001	trivial
V2_WCF_Q_8	0	trivial
V2_WCF_Q_9	0.004	trivial

DIF by Race: Latino/Hispanic vs. Asian

Item	Pseudo-r^2 difference	DIF effect size

V2_WCF_Q_1	-0.002	trivial
V2_WCF_Q_10	0.001	trivial
V2_WCF_Q_11	0.002	trivial
V2_WCF_Q_2	0	trivial
V2_WCF_Q_3	-0.004	trivial
V2_WCF_Q_4	0.002	trivial
V2_WCF_Q_5	0	trivial
V2_WCF_Q_6	-0.003	trivial
V2_WCF_Q_7	0.001	trivial
V2_WCF_Q_8	0.001	trivial
V2_WCF_Q_9	-0.002	trivial

RL Scale

DIF by Gender

Item	Pseudo-r^2 difference	DIF effect size
V2_RL_Q_1	0.002	trivial
V2_RL_Q_10	-0.005	trivial
V2_RL_Q_11	-0.003	trivial
V2_RL_Q_2	-0.002	trivial
V2_RL_Q_3	-0.004	trivial
V2_RL_Q_4	0	trivial
V2_RL_Q_5	0.007	trivial
V2_RL_Q_6	-0.003	trivial
V2_RL_Q_7	-0.003	trivial
V2_RL_Q_8	0.002	trivial
V2_RL_Q_9	-0.009	trivial

DIF by Race: Latino/Hispanic vs. Black

Item	Pseudo-r^2 difference	DIF effect size
V2_RL_Q_1	0.011*	trivial
V2_RL_Q_10	-0.007	trivial
V2_RL_Q_11	-0.001	trivial
V2_RL_Q_2	0	trivial
V2_RL_Q_3	0.004	trivial
V2_RL_Q_4	0.024	trivial

V2_RL_Q_5	0.001	trivial
V2_RL_Q_6	-0.002*	trivial
V2_RL_Q_7	-0.012	trivial
V2_RL_Q_8	0.004	trivial
V2_RL_Q_9	-0.004*	trivial

DIF by Race: Latino/Hispanic vs. White

Item	Pseudo-r^2 difference	DIF effect size
V2_RL_Q_1	-0.002	trivial
V2_RL_Q_10	0.002	trivial
V2_RL_Q_11	-0.006	trivial
V2_RL_Q_2	0.004	trivial
V2_RL_Q_3	-0.009	trivial
V2_RL_Q_4	0.002	trivial
V2_RL_Q_5	0	trivial
V2_RL_Q_6	0.002	trivial
V2_RL_Q_7	-0.002	trivial
V2_RL_Q_8	0	trivial
V2_RL_Q_9	0.007	trivial

DIF by Race: Latino/Hispanic vs. Asian

Item	Pseudo-r^2 difference	DIF effect size
V2_RL_Q_1	-0.002	trivial
V2_RL_Q_10	0	trivial
V2_RL_Q_11	-0.004	trivial
V2_RL_Q_2	0.004	trivial
V2_RL_Q_3	0.004*	trivial
V2_RL_Q_4	-0.006	trivial
V2_RL_Q_5	0	trivial
V2_RL_Q_6	0	trivial
V2_RL_Q_7	0	trivial
V2_RL_Q_8	0	trivial
V2_RL_Q_9	0.01	trivial

REL Scale

DIF by Gender

Item	Pseudo-r^2 difference	DIF effect size
V2_REL_Q_1	0	trivial
V2_REL_Q_10	0.002	trivial
V2_REL_Q_11	0.002	trivial
V2_REL_Q_2	-0.002*	trivial
V2_REL_Q_3	0.011*	trivial
V2_REL_Q_4	0.002	trivial
V2_REL_Q_5	0	trivial
V2_REL_Q_6	0.012	trivial
V2_REL_Q_7	0.001	trivial
V2_REL_Q_8	0.005	trivial
V2_REL_Q_9	-0.005	trivial

DIF by Race: Latino/Hispanic vs. Black

Item	Pseudo-r^2 difference	DIF effect size
V2_REL_Q_1	0	trivial
V2_REL_Q_10	-0.003	trivial
V2_REL_Q_11	0.01	trivial
V2_REL_Q_2	0.008	trivial
V2_REL_Q_3	0.008	trivial
V2_REL_Q_4	-0.008	trivial
V2_REL_Q_5	-0.014	trivial
V2_REL_Q_6	0.002	trivial
V2_REL_Q_7	-0.014	trivial
V2_REL_Q_8	0.008	trivial
V2_REL_Q_9	0.002	trivial

DIF by Race: Latino/Hispanic vs. White

Item	Pseudo-r^2 difference	DIF effect size
V2_REL_Q_1	0	trivial
V2_REL_Q_10	0	trivial
V2_REL_Q_11	0.003	trivial
V2_REL_Q_2	-0.009*	trivial

V2_REL_Q_3	-0.002	trivial
V2_REL_Q_4	0.002	trivial
V2_REL_Q_5	0.005	trivial
V2_REL_Q_6	0.004	trivial
V2_REL_Q_7	-0.012*	trivial
V2_REL_Q_8	0	trivial
V2_REL_Q_9	0.002	trivial

DIF by Race: Latino/Hispanic vs. Asian

Item	Pseudo-r^2 difference	DIF effect size
V2_REL_Q_1	0	trivial
V2_REL_Q_10	0.008*	trivial
V2_REL_Q_11	0	trivial
V2_REL_Q_2	0	trivial
V2_REL_Q_3	0.004	trivial
V2_REL_Q_4	-0.002	trivial
V2_REL_Q_5	0.01	trivial
V2_REL_Q_6	-0.002	trivial
V2_REL_Q_7	-0.009	trivial
V2_REL_Q_8	0.004	trivial
V2_REL_Q_9	0.004	trivial

ASO Scale

DIF by Gender

Item	Pseudo-r^2 difference	DIF effect size
V2_ASO_Q_1	-0.001*	trivial
V2_ASO_Q_10	0	trivial
V2_ASO_Q_11	0	trivial
V2_ASO_Q_2	0	trivial
V2_ASO_Q_3	0	trivial
V2_ASO_Q_4	0*	trivial
V2_ASO_Q_5	-0.001	trivial
V2_ASO_Q_6	0	trivial
V2_ASO_Q_7	0.002	trivial
V2_ASO_Q_8	0	trivial

DIF by Race: Latino/Hispanic vs. Black

Item	Pseudo-r^2 difference	DIF effect size
V2_ASO_Q_1	0	trivial
V2_ASO_Q_10	0.001	trivial
V2_ASO_Q_11	0.001	trivial
V2_ASO_Q_2	0.004*	trivial
V2_ASO_Q_3	-0.004*	trivial
V2_ASO_Q_4	0*	trivial
V2_ASO_Q_5	0	trivial
V2_ASO_Q_6	0	trivial
V2_ASO_Q_7	-0.005*	trivial
V2_ASO_Q_8	0.003*	trivial

DIF by Race: Latino/Hispanic vs. White

Item	Pseudo-r^2 difference	DIF effect size
V2_ASO_Q_1	0	trivial
V2_ASO_Q_10	0.003*	trivial
V2_ASO_Q_11	0	trivial
V2_ASO_Q_2	-0.001	trivial
V2_ASO_Q_3	0.002	trivial
V2_ASO_Q_4	-0.001	trivial
V2_ASO_Q_5	0.005*	trivial
V2_ASO_Q_6	0.009	trivial
V2_ASO_Q_7	0	trivial
V2_ASO_Q_8	0.002	trivial

DIF by Race: Latino/Hispanic vs. Asian

Item	Pseudo-r^2 difference	DIF effect size
V2_ASO_Q_1	0	trivial
V2_ASO_Q_10	-0.004*	trivial
V2_ASO_Q_11	0.003*	trivial
V2_ASO_Q_2	0.006*	trivial
V2_ASO_Q_3	0.001	trivial
V2_ASO_Q_4	0.001*	trivial
V2_ASO_Q_5	0	trivial

V2_ASO_Q_6	0.002	trivial
V2_ASO_Q_7	0.001	trivial
V2_ASO_Q_8	0.001*	trivial

SCA Scale

DIF by Gender

Item	Pseudo-r^2 difference	DIF effect size
V2_SCA_Q_13	-0.003*	trivial
V2_SCA_Q_16	0	trivial
V2_SCA_Q_17	-0.009	trivial
V2_SCA_Q_2	0.001	trivial
V2_SCA_Q_3	0	trivial
V2_SCA_Q_4	-0.012	trivial
V2_SCA_Q_5	0.014	trivial
V2_SCA_Q_6	0.001	trivial
V2_SCA_Q_7	0	trivial
V2_SCA_Q_8	0.007	trivial
V2_SCA_Q_9	-0.003	trivial

DIF by Race: Latino/Hispanic vs. Black

Item	Pseudo-r^2 difference	DIF effect size
V2_SCA_Q_13	0.002	trivial
V2_SCA_Q_16	0.021	trivial
V2_SCA_Q_17	0	trivial
V2_SCA_Q_2	0.022*	trivial
V2_SCA_Q_3	0	trivial
V2_SCA_Q_4	-0.007	trivial
V2_SCA_Q_5	0.005	trivial
V2_SCA_Q_6	-0.008	trivial
V2_SCA_Q_7	-0.003	trivial
V2_SCA_Q_8	0.002	trivial
V2_SCA_Q_9	0.002	trivial

DIF by Race: Latino/Hispanic vs. White

Item	Pseudo-r^2 difference	DIF effect size

V2_SCA_Q_13	-0.004	trivial
V2_SCA_Q_16	0	trivial
V2_SCA_Q_17	0	trivial
V2_SCA_Q_2	0.01*	trivial
V2_SCA_Q_3	0.002	trivial
V2_SCA_Q_4	-0.017	trivial
V2_SCA_Q_5	0	trivial
V2_SCA_Q_6	0.002	trivial
V2_SCA_Q_7	0.002	trivial
V2_SCA_Q_8	0	trivial
V2_SCA_Q_9	0.004*	trivial

DIF by Race: Latino/Hispanic vs. Asian

Item	Pseudo-r^2 difference	DIF effect size
V2_SCA_Q_13	0.002	trivial
V2_SCA_Q_16	-0.002	trivial
V2_SCA_Q_17	0.004	trivial
V2_SCA_Q_2	0	trivial
V2_SCA_Q_3	0	trivial
V2_SCA_Q_4	-0.013*	trivial
V2_SCA_Q_5	0	trivial
V2_SCA_Q_6	0	trivial
V2_SCA_Q_7	-0.004	trivial
V2_SCA_Q_8	0	trivial
V2_SCA_Q_9	-0.006	trivial

CC Scale

DIF by Gender

Item	Pseudo-r^2 difference	DIF effect size
V2_CC_Q_1	-0.004*	trivial
V2_CC_Q_10	0	trivial
V2_CC_Q_2	-0.004*	trivial
V2_CC_Q_3	0.003*	trivial
V2_CC_Q_4	-0.002*	trivial
V2_CC_Q_5	0	trivial

V2_CC_Q_6	0.004*	trivial
V2_CC_Q_7	-0.002*	trivial
V2_CC_Q_8	0*	trivial
V2_CC_Q_9	0.003*	trivial

DIF by Race: Latino/Hispanic vs. Black

Item	Pseudo-r^2 difference	DIF effect size
V2_CC_Q_1	0.003*	trivial
V2_CC_Q_10	0	trivial
V2_CC_Q_2	-0.005*	trivial
V2_CC_Q_3	0.005	trivial
V2_CC_Q_4	0	trivial
V2_CC_Q_5	0*	trivial
V2_CC_Q_6	0	trivial
V2_CC_Q_7	-0.007	trivial
V2_CC_Q_8	-0.002	trivial
V2_CC_Q_9	0.002*	trivial

DIF by Race: Latino/Hispanic vs. White

Item	Pseudo-r^2 difference	DIF effect size
V2_CC_Q_1	0.001*	trivial
V2_CC_Q_10	-0.003	trivial
V2_CC_Q_2	-0.008*	trivial
V2_CC_Q_3	-0.007*	trivial
V2_CC_Q_4	0*	trivial
V2_CC_Q_5	0	trivial
V2_CC_Q_6	0.002	trivial
V2_CC_Q_7	-0.002	trivial
V2_CC_Q_8	-0.002	trivial
V2_CC_Q_9	0.005*	trivial

DIF by Race: Latino/Hispanic vs. Asian

Item	Pseudo-r^2 difference	DIF effect size
V2_CC_Q_1	0	trivial
V2_CC_Q_10	0	trivial

V2_CC_Q_2	0.001	trivial
V2_CC_Q_3	0	trivial
V2_CC_Q_4	0.002	trivial
V2_CC_Q_5	0.001	trivial
V2_CC_Q_6	0.003	trivial
V2_CC_Q_7	-0.001	trivial
V2_CC_Q_8	0.003	trivial
V2_CC_Q_9	0.002	trivial

CUS Scale

DIF by Gender

Item	Pseudo-r^2 difference	DIF effect size
V2_CUS_Q_1	-0.006	trivial
V2_CUS_Q_10	0.002	trivial
V2_CUS_Q_11	0.004	trivial
V2_CUS_Q_12	-0.003	trivial
V2_CUS_Q_13	0.003	trivial
V2_CUS_Q_14	0.014	trivial
V2_CUS_Q_15	0.002	trivial
V2_CUS_Q_2	0*	trivial
V2_CUS_Q_3	-0.002	trivial
V2_CUS_Q_4	-0.005	trivial
V2_CUS_Q_5	-0.002	trivial
V2_CUS_Q_6	0.009	trivial
V2_CUS_Q_7	0.008	trivial
V2_CUS_Q_8	0.002	trivial
V2_CUS_Q_9	-0.005	trivial

DIF by Race: Latino/Hispanic vs. Black

Item	Pseudo-r^2 difference	DIF effect size
V2_CUS_Q_1	-0.003*	trivial
V2_CUS_Q_10	0	trivial
V2_CUS_Q_11	0.001*	trivial
V2_CUS_Q_12	-0.001	trivial
V2_CUS_Q_13	-0.001	trivial

V2_CUS_Q_14	-0.001	trivial
V2_CUS_Q_15	0.016*	trivial
V2_CUS_Q_2	-0.003	trivial
V2_CUS_Q_3	0.004*	trivial
V2_CUS_Q_4	-0.004	trivial
V2_CUS_Q_5	-0.001	trivial
V2_CUS_Q_6	-0.006	trivial
V2_CUS_Q_7	0.003	trivial
V2_CUS_Q_8	0.009	trivial
V2_CUS_Q_9	-0.003	trivial

DIF by Race: Latino/Hispanic vs. White

Item	Pseudo-r^2 difference	DIF effect size
V2_CUS_Q_1	0.009	trivial
V2_CUS_Q_10	0.003	trivial
V2_CUS_Q_11	-0.004	trivial
V2_CUS_Q_12	-0.002	trivial
V2_CUS_Q_13	0.01	trivial
V2_CUS_Q_14	0	trivial
V2_CUS_Q_15	0.002	trivial
V2_CUS_Q_2	0.005	trivial
V2_CUS_Q_3	0*	trivial
V2_CUS_Q_4	0.009	trivial
V2_CUS_Q_5	0	trivial
V2_CUS_Q_6	0.002	trivial
V2_CUS_Q_7	0.002	trivial
V2_CUS_Q_8	0.019	trivial
V2_CUS_Q_9	0.008	trivial

DIF by Race: Latino/Hispanic vs. Asian

Item	Pseudo-r^2 difference	DIF effect size
V2_CUS_Q_1	0	trivial
V2_CUS_Q_10	0.002	trivial
V2_CUS_Q_11	0	trivial
V2_CUS_Q_12	-0.002	trivial

V2_CUS_Q_13	0	trivial
V2_CUS_Q_14	0.002	trivial
V2_CUS_Q_15	-0.002	trivial
V2_CUS_Q_2	0.004	trivial
V2_CUS_Q_3	0.006	trivial
V2_CUS_Q_4	0	trivial
V2_CUS_Q_5	0.002	trivial
V2_CUS_Q_6	0	trivial
V2_CUS_Q_7	0.01	trivial
V2_CUS_Q_8	0.002	trivial
V2_CUS_Q_9	0.002	trivial

ASD Scale

DIF by Gender

Item	Pseudo-r^2 difference	DIF effect size
V2_ASD_Q_1	-0.042*	trivial
V2_ASD_Q_2	0.006	trivial
V2_ASD_Q_3	0	trivial
V2_ASD_Q_4	0	trivial
V2_ASD_Q_5	-0.007	trivial
V2_ASD_Q_6	0.006	trivial
V2_ASD_Q_7	0.003	trivial
V2_ASD_Q_8	0.005*	trivial
V2_ASD_Q_9	-0.002	trivial

DIF by Race: Latino/Hispanic vs. Black

Item	Pseudo-r^2 difference	DIF effect size
V2_ASD_Q_1	-0.025	trivial
V2_ASD_Q_2	0	trivial
V2_ASD_Q_3	0	trivial
V2_ASD_Q_4	0.003	trivial
V2_ASD_Q_5	-0.002	trivial
V2_ASD_Q_6	-0.012*	trivial
V2_ASD_Q_7	0	trivial
V2_ASD_Q_8	0	trivial

V2_ASD_Q_9	0.007	trivial

DIF by Race: Latino/Hispanic vs. White

Item	Pseudo-r^2 difference	DIF effect size
V2_ASD_Q_1	0.011	trivial
V2_ASD_Q_2	0.007*	trivial
V2_ASD_Q_3	0.005	trivial
V2_ASD_Q_4	0.002	trivial
V2_ASD_Q_5	0.012*	trivial
V2_ASD_Q_6	0.003*	trivial
V2_ASD_Q_7	0	trivial
V2_ASD_Q_8	-0.017	trivial
V2_ASD_Q_9	0.007	trivial

DIF by Race: Latino/Hispanic vs. Asian

Item	Pseudo-r^2 difference	DIF effect size
V2_ASD_Q_1	-0.02*	trivial
V2_ASD_Q_2	0.009	trivial
V2_ASD_Q_3	0	trivial
V2_ASD_Q_4	0	trivial
V2_ASD_Q_5	0.006	trivial
V2_ASD_Q_6	-0.012	trivial
V2_ASD_Q_7	0.003	trivial
V2_ASD_Q_8	-0.001	trivial
V2_ASD_Q_9	0	trivial

AAL Scale

DIF by Gender

Item	Pseudo-r^2 difference	DIF effect size
V2_AAL_Q_1	-0.003*	trivial
V2_AAL_Q_10	0.002*	trivial
V2_AAL_Q_11	-0.006	trivial
V2_AAL_Q_12	0.011	trivial
V2_AAL_Q_13	-0.031*	trivial
V2_AAL_Q_14	-0.009	trivial

V2_AAL_Q_15	0.004	trivial
V2_AAL_Q_16	-0.023*	trivial
V2_AAL_Q_17	0.006	trivial
V2_AAL_Q_2	-0.007	trivial
V2_AAL_Q_3	0.004	trivial
V2_AAL_Q_4	0.005	trivial
V2_AAL_Q_5	0.004	trivial
V2_AAL_Q_6	-0.007	trivial
V2_AAL_Q_7	-0.016	trivial
V2_AAL_Q_8	0.004	trivial
V2_AAL_Q_9	-0.014	trivial

DIF by Race: Latino/Hispanic vs. Black

Item	Pseudo-r^2 difference	DIF effect size
V2_AAL_Q_1	0.008	trivial
V2_AAL_Q_10	-0.011*	trivial
V2_AAL_Q_11	0.003	trivial
V2_AAL_Q_12	-0.01	trivial
V2_AAL_Q_13	-0.011	trivial
V2_AAL_Q_14	0.003	trivial
V2_AAL_Q_15	0.005	trivial
V2_AAL_Q_16	-0.024*	trivial
V2_AAL_Q_17	0.002	trivial
V2_AAL_Q_2	0.012	trivial
V2_AAL_Q_3	0	trivial
V2_AAL_Q_4	-0.006	trivial
V2_AAL_Q_5	0.01	trivial
V2_AAL_Q_6	0	trivial
V2_AAL_Q_7	0.01*	trivial
V2_AAL_Q_8	-0.002	trivial
V2_AAL_Q_9	0.002	trivial

DIF by Race: Latino/Hispanic vs. White

Item	Pseudo-r^2 difference	DIF effect size
V2_AAL_Q_1	0.011	trivial

V2_AAL_Q_10	-0.002	trivial
V2_AAL_Q_11	0	trivial
V2_AAL_Q_12	0.004	trivial
V2_AAL_Q_13	-0.002	trivial
V2_AAL_Q_14	0	trivial
V2_AAL_Q_15	0.002	trivial
V2_AAL_Q_16	0.008	trivial
V2_AAL_Q_17	0.006	trivial
V2_AAL_Q_2	-0.005	trivial
V2_AAL_Q_3	0.007*	trivial
V2_AAL_Q_4	0	trivial
V2_AAL_Q_5	0	trivial
V2_AAL_Q_6	0.004	trivial
V2_AAL_Q_7	0.012	trivial
V2_AAL_Q_8	0.008*	trivial
V2_AAL_Q_9	-0.002	trivial

DIF by Race: Latino/Hispanic vs. Asian

Item	Pseudo-r^2 difference	DIF effect size
V2_AAL_Q_1	0.002	trivial
V2_AAL_Q_10	0.002	trivial
V2_AAL_Q_11	0.004	trivial
V2_AAL_Q_12	-0.011	trivial
V2_AAL_Q_13	0.008	trivial
V2_AAL_Q_14	0.004	trivial
V2_AAL_Q_15	0.004	trivial
V2_AAL_Q_16	0.002	trivial
V2_AAL_Q_17	-0.004	trivial
V2_AAL_Q_2	0.008	trivial
V2_AAL_Q_3	0	trivial
V2_AAL_Q_4	-0.005	trivial
V2_AAL_Q_5	0*	trivial
V2_AAL_Q_6	0.006	trivial
V2_AAL_Q_7	-0.002	trivial
V2_AAL_Q_8	0	trivial

V2_AAL_Q_9	0.002	trivial

Appendix C. Full Text for Leaps Pulse Check Scale Items

Scale	ZItem	Item Text
HE	V2_HE_Q_1	At my school, it feels like I'm expected -- and supported -- to learn a ton.
	V2_HE_Q_2	At my school, people don't give up when the work gets hard.
	V2_HE_Q_3	I feel like I have access to all of the opportunities my school offers.
	V2_HE_Q_13	When you feel like giving up on a difficult task, how likely is it that someone at your school will help you keep trying?
WCF	V2_WCF_Q_1	At my school I learn to understand my emotions.
	V2_WCF_Q_2	At my school I learn how to figure out who I am as a person.
	V2_WCF_Q_11	At my school, everyone wants me to not just learn, but also be happy and healthy in mind and body.
RL	V2_RL_Q_1	At my school we use our thinking skills, rather than just memorizing things.
	V2_RL_Q_2	At my school I get to develop my own ideas.
	V2_RL_Q_7	At my school we have time to explain our ideas.
REL	V2_REL_Q_1	At my school what we learn is often connected to life outside the classroom.
	V2_REL_Q_2	At my school I get to learn things I'm interested in.
	V2_REL_Q_3	At my school what I'm learning matters a lot to me.
ASO	V2_ASO_Q_2	At my school I feel proud of who I am.
	V2_ASO_Q_3	I can be myself at my school.

	V2_ASO_Q_4	At my school it feels like being yourself is a great thing. I feel safe and appreciated for who I am.
SCA	V2_SCA_Q_16	At my school we learn about racism in the United States.
	V2_SCA_Q_17	At my school we take action to fix problems in society, such as racism and discrimination.
CC	V2_CC_Q_1	I feel part of the community at my school. There are a lot of people who know and care about me.
	V2_CC_Q_2	Overall, how much do you feel like you belong at your school?
	V2_CC_Q_4	At my school I feel included by other students.
CUS	V2_CUS_Q_2	At my school I do work that meets me where I am in my learning.
	V2_CUS_Q_8	At my school I have the resources I need to support my learning.
	V2_CUS_Q_13	At my school I am able to catch up if I am behind.
ASD	V2_ASD_Q_1	At my school I have goals for my learning, and I have choices about how I pursue those goals.
	V2_ASD_Q_2	At my school I feel like I have a say about what happens to me.
	V2_ASD_Q_3	At my school I can choose how to do my work.
	V2_ASD_Q_6	Adults at my school respect my ideas and suggestions.
AAL	V2_AAL_Q_1	At my school we spend time learning outside of our school building.
	V2_AAL_Q_2	I interact with people outside my school to help me learn.

	V2_AAL_Q_15	I can make progress on my learning when I am not at my school.
	V2_AAL_Q_17	My school encourages and gives me the tools to be learning everywhere, from everyone, all the time.

Note. References to “my school”, “your school”, or “school” could be replaced with actual organization names in practice.