Validity Study

Social & Emotional Learning Indicator System (SELIS) Project

Student Report (2021)

VO

### 

|  |
| --- |
| ESE logo  This document was prepared by the  Massachusetts Department of Elementary and Secondary Education  Jeffrey Riley  Commissioner  The Massachusetts Department of Elementary and Secondary Education, an affirmative action employer, is committed to ensuring that all of its programs and facilities are accessible to all members of the public.  We do not discriminate on the basis of age, color, disability, national origin, race, religion, sex, gender identity, or sexual orientation.  Inquiries regarding the Department’s compliance with Title IX and other civil rights laws may be directed to the  Human Resources Director, 75 Pleasant St., Malden, MA 02148-4906. Phone: 781-338-6105.  © 2020 Massachusetts Department of Elementary and Secondary Education  Permission is hereby granted to copy any or all parts of this document for non-commercial educational purposes. Please credit the “Massachusetts Department of Elementary and Secondary Education.”  This document printed on recycled paper  Massachusetts Department of Elementary and Secondary Education  75 Pleasant Street, Malden, MA 02148-4906  Phone 781-338-3000 TTY: N.E.T. Relay 800-439-2370  www.doe.mass.edu  State Seal of Massachusetts |

Table of Contents

Purpose and design of this report…………………………………………………………......…......... 1

Pilot participants (profile of respondents)…………………………………....................................... 1

Survey content and design …………………………………............................................................ 2

Survey administration and response rates…….……..................................................................... 3

Validity framework ……………………………………………………………….................................. 4

Content validity……………………………………………………...……................................. 5

Structural validity …………………………………………………………................................ 5

Substantive validity………………………………………………………................................. 10

Generalizability………………………………………………………..….................................. 20

External validity…………………………………………………………................................... 30

Consequential validity………………………………………………………............................. 36

Conclusion and limitations of study………………………………...…………................................. 36 References…………………………………………………………………………............................... 38

Appendices

1. Collaborative for Academic and Social and Emotional Learning (CASEL) conceptual framework.................................................................................................................... 39

B. SELIS survey item prompts ……………………….............................................…........40

C. SELIS survey specifications....................................... ……………………………......… 42

D. Intra-class correlations for SELIS data..............................……………………….......… 43

E. Linear regression models (predicting SELIS scores with achievement variables)....... 44

Purpose and Design of this Report

This report offers reliability and validity evidence to support the use of the Social and Emotional Learning Indicator System (SELIS) student survey piloted by the Massachusetts Department of Elementary and Secondary Education (DESE). DESE administered the SELIS survey during the 2020-2021 academic year to: (1) differentiate levels of students’ social and emotional competencies within and between schools, and (2) provide schools and districts with concrete, actionable information about students’ social and emotional learning in order to inform a tiered system of support for students.

This report is intended for readers with knowledge of survey development and validation, psychometrics, and educational measurement. Familiarity with Wolfe and Smith’s (2007a, 2007b) and Messick’s (1995a) construct validity frameworks for instrument development is helpful. Social and emotional learning (SE) is a psychological construct; students provide their perceptions of their SE skills by responding to statements in the SELIS survey. Evidence from six aspects of construct validity (i.e., content, structural, substantive, generalizability, external, and consequential) combine to justify the use of SELIS scores as a measure of students’ perceptions of their SE skills. All six validity aspects are addressed in this study; coverage of external and consequential validity is relatively limited when compared to the other four aspects.

The report is predominantly graphical and tabular with text used to briefly summarize the findings.

Pilot Participants (Profile of Respondents)

Ten districts from across Massachusetts participated in the pilot. Three districts are classified as urban, 2 districts are considered rural, with the remaining districts deemed suburban. The districts represent all five regions of Massachusetts. Eighty-two schools from across the 10 districts participated in the SELIS project. Overall, 16,039 students provided data for this validity study. The profile of respondents is shown in Table 1. The percentage of each student group across the 10 districts is reasonably representative of the state; however, white students are slightly over-represented in the pilot sample. Five of the ten districts only had students from grades 4, 5, 8, and 10 participate; as a result, these grades are over-represented in the pilot sample.

Table 1: Profile of respondents

| **Pilot Group** | **Student Group** | **Sample**  **Number1** | **Sample Percent** | **State (Percent)** |
| --- | --- | --- | --- | --- |
| Gender | Female | 8082 | 49.3 | 48.7 |
| Male | 7949 | 50.7 | 51.3 |
| Non-binary | ---- | <0.01 | <0.1 |
| Race/ethnicity of students | Asian | 762 | 6.6 | 7.2 |
| Black | 948 | 9.3 | 9.3 |
| Hispanic | 3642 | 21.2 | 22.3 |
| Other2 | 718 | 4.0 | 4.4 |
| White | 9968 | 59.0 | 56.7 |
| Other student groups | Economically disadvantaged | 6059 | 36.7 | 36.6 |
| English learner | 1156 | 9.0 | 10.5 |
| Students with disabilities | 2630 | 17.9 | 18.5 |
| Grade3 | Grade 3 | 1068 | 6.7 |  |
| Grade 4 | 3173 | 19.8 |  |
| Grade 5 | 3294 | 20.5 |  |
| Grade 6 | 1198 | 7.5 |  |
| Grade 7 | 1044 | 6.5 |  |
| Grade 8 | 2954 | 18.4 |  |
| Grade 9 | 653 | 4.1 |  |
| Grade 10 | 1937 | 12.1 |  |
| Grade 11 | 568 | 3.5 |  |
| Grade 12 | 150 | 0.9 |  |
| Total | All students | 16039 | ---- | ------ |

1Numbers may not add to total due to missing demographic information; 2Includes Multi-race, Non-Hispanic, Native American, and Native Hawaiian, Pacific Islander students; 3Five of the ten districts only had students from grades 4, 5, 8, and 10 participate. Dashed lines indicate suppression due to small sample size.

SELIS survey content and survey design

SELIS content. The SELIS survey items are designed to measure students’ social and emotional (SE) skills in five dimensions. These dimensions were founded on the Collaborative for Social and Emotional and Academic Learning’s (CASEL) conceptual framework (Appendix A). The five dimensions are students’ self-awareness skills, students’ self-management skills, students’ relationship skills, students’ social awareness skills, and students’ responsible decision-making skills. The dimensions are defined as follows:

1. **Self-awareness skills (SA):** This dimension measures students’ abilities to understand one’s own emotions, thoughts, and values and how they influence behavior across contexts.
2. **Self-management skills (SM):** This dimension measures students’ abilities to manage one’s emotions, thoughts, and behaviors effectively in different situations and to achieve goals and aspirations.
3. **Social awareness skills (SOC):** This dimension measures students’ abilities to understand the perspectives of and empathize with others, including those with different backgrounds, cultures, and contexts than their own.
4. **Relationship skills (RSK):** This dimension measures students’ abilities to establish and maintain healthy and supportive relationships and to effectively navigate diverse settings.
5. **Responsible decision-making skills (RDM):** This dimension measures students’ abilities to make caring and constructive choices about personal behavior and social interactions across diverse situations.

The SELIS survey is predicated on the Social and Emotional Competency Assessment developed by a team of researchers, educational practitioners and CASEL[[1]](#footnote-2). New items were also developed to try to improve the responsiveness of each of the five dimensions (Appendix B). Students respond, “Very Easy” (scored 3), “Easy” (scored 2), “Hard” (scored 1), or “Very hard” (scored 0) to the skills statements.

SELIS design. Appendix C shows the specifications for the survey administration. Students in grades 3 to 5 responded to 45 items in total; students in grades 6 to 8 and in grades 9 to 11 responded to 48 and 50 items, respectively. Forty-one items were common across all grades and these items were used to anchor all 61 items to the same scale metric. The remaining 20 items were distributed across the three grade-level forms and were unique to each form. As a result of the anchoring process, all items and all students are measured on the same scale metric.

The survey was translated into four languages: Brazilian Portuguese, Haitian Creole, Simplified Chinese, and Spanish. Students were able to choose their preferred language. Almost ninety-nine percent of surveys were taken in English (see Table 2).

Survey administration and response rates

A computer-based survey administration was used. A dropdown was used by students to choose their preferred language. Students watched a two-minute video explaining how to navigate the survey platform. The survey was read out loud to students in grade 3. Table 2 provides data on student response rates. Overall, 66% of students in the districts’ rosters provided usable surveys; the percentage of usable surveys at the district-level ranged from 42% to 89%.

Table 2: Response rates and language of survey data

| Type of Data | Student Group  (percent1) | Sample |
| --- | --- | --- |
| Response Rates Data | Number of students in district rosters | 24,295 |
| Number of students participating1 | 18,485 |
| Usable surveys2 | 16,039 |
| Percent usable/roster | 66% |
| Percent usable/participating | 87% |
| Range of districts’ students participating (percent) | 51% – 98% |
| Range of districts’ students with usable surveys (percent) | 42% – 89% |
| Language used to take the survey | Percent of usable student surveys taken in Brazilian Portuguese | 0.6 |
| Percent of usable student surveys taken in English | 98.9 |
| Percent of usable student surveys taken in Haitian Creole | <0.01 |
| Percent of usable student surveys taken in Simplified Chinese | <0.01 |
| Percent of usable student surveys taken in Spanish | 0.5 |

1Students participating are the number of students who clicked on the survey link; 2A usable survey is one where a student responded to 20 or more items in the survey.

Validity framework

Messick’s (1980, 1995a) unified concept of construct validity guided the validity analyses for the SE construct. Messick (1995a, p. 741) defines validity as “an evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessment.” Table 3 outlines the specific validity aspects addressed in this validity study.

Wolfe and Smith (2007a, 2007b, p. 205) used Messick’s validity conceptualization to detail instrument development activities and evidence that are needed to support the use of scores from instruments based on the Rasch measurement framework and these are illuminated in Table 2. This report primarily focuses on internal validity with more limited external validity evidence provided for the social and emotional learning construct.

Table 3: Rasch-Based instrument validity framework and evidence collected for SELIS survey1

| Validity evidence | | |
| --- | --- | --- |
| Content | Structural | Substantive |
| Instrument purpose  Test specification  Item technical quality | Principal components residual analyses  Rasch dimensionality analyses  Sub-scale correlations | Rating scale functioning  Item difficulty hierarchy |
| Validity evidence | | |
| Generalizability | External | Consequential2 |
| Differential item functioning  Person separation reliability  Item invariance | Responsiveness  Relationship between SELIS scaled-scores with scores from similar/dissimilar constructs | *Standard setting*  Score use |

1 Validity framework is based on: Messick (1995a) and Wolfe and Smith (2007b) conceptualization and representation. 2Standard setting is not part of this study.

Content Validity

Content validity examines the “content relevance, representativeness and technical quality” (Messick, 1995a, p.745) of the items used as indicators of the construct. The SELIS content has been described in a previous section; this section focuses on the technical quality of the items. The outfit mean square error (OMNSQ) fit statistic was used in this study to assess item technical quality; this statistic provides the most stable fit statistic and is least impacted by large sample sizes (Smith, 2008). Item OMNSQ statistics of between 0.5 and 1.5 are productive for measurement (Wright and Linacre, 1994; Boone, Staver, and Yale, 2014; Linacre, 2021). In addition, the point-to-measure (PTM) correlation was included in the assessment. PTMs below 0.3 indicate that the item is not likely construct relevant.

When all 61 SELIS items were calibrated together, all item OMNSQ fit statistics fall between 0.50 and 1.5 indicating they are largely well-fitting (Table 4). One item’s fit statistic, RDM10, is distinct with an OMNSQ of 0.57; this item should be reviewed for construct redundancy and possible omission from future administrations. PTMs range from 0.32 (SOC1) to 0.62 (SME13) indicating that all items are construct relevant; however, items with PTMs below 0.4 will be investigated to determine if they can be improved for future administrations.

Structural Validity

Structural validity evaluates the alignment of the scoring structure to the hypothesized structure of the construct. The fundamental assumption of the Rasch model is that it measures only one latent unidimensional construct (in this study, the social and emotional learning construct). The unidimensionality of the data was assessed by conducting (1) an assessment of the dimensionality data provided by the Rasch Winsteps software (Linacre, 2021), (2) an analysis of the standardized residuals, and (3) an examination of the correlational relationship between the freely calibrated dimension scores. These analyses were done for all 61 SELIS items and separately for items belonging to each of the five dimensions.

The variance explained by the 61-item measure was 34.1%. The Eigenvalue of the first contrast was 3.3 and represents 3.6% of the observed variance (<5% supports unidimensionality). However, the variance explained by the items of the first dimension (SEL construct) does not meet the criterion of being 4 times the explained variance of the 1st contrast (Linacre, 2021); the multiple is 2.75x.

Four items from the self-management dimension make up the 1st contrast (SMS3, SMS9, SMS10, SMG11); these items’ content are related as they all measure some facet of “perseverance” and have broken out to form the 1st contrast. Theoretically, perseverance is an essential facet of students’ ability to self-manage.

The standardized residual plot (Figure 1) divides the items into 3 clusters; items in cluster 1 and cluster 3 are most different. If the disattenuated correlations between the person measures for the three clusters are below 0.7, this provides evidence that the construct is not unidimensional. The disattenuated correlation between Cluster 1 (which contains the four items of the 1st contrast) and Cluster 3 is 0.75 which indicates that the items forming the two clusters share more than half their variance and are dependent. The magnitude of the cluster correlations and the premise that “perseverance” is an essential theoretical component of the SE construct provides sufficient evidence to support the unidimensionality of the SE construct.

To make the data more actionable for schools and districts, educators were provided with scaled scores for each of the five dimensions of SE construct. Table 5 provides a summary of the dimensionality data for each dimension. Each dimension, when calibrated separately, forms a unidimensional construct.

-------------------------------------------------------------------------------------------

Table 4: Item statistics: Misfit Order

-------------------------------------------------------------------------------------------

|ENTRY TOTAL TOTAL MODEL| INFIT | OUTFIT |PTMEASUR-AL|EXACT MATCH| |

|NUMBER SCORE COUNT MEASURE S.E. |MNSQ ZSTD|MNSQ ZSTD|CORR. EXP.| OBS% EXP%| ITEM |

|------------------------------------+----------+----------+-----------+-----------+------|

| 38 14982 7240 -.49 .02|1.41 9.90|1.39 9.90|A .45 .51| 46.2 54.0| SOC6 |

| 29 28036 15440 .07 .01|1.40 9.90|1.40 9.90|B .47 .51| 45.4 51.5| SME15|

| 3 32126 15778 -.38 .01|1.35 9.90|1.38 9.90|C .35 .48| 47.6 54.0| SAE3 |

| 55 34289 15805 -.68 .01|1.37 9.90|1.36 9.90|D .41 .46| 49.8 55.1| RDM4 |

| 33 30829 15828 -.20 .01|1.29 9.90|1.35 9.90|E .32 .49| 48.4 53.0| SOC1 |

| 12 16172 7268 -.85 .02|1.33 9.90|1.27 9.90|F .47 .49| 51.1 55.7| SAS12|

| 16 25830 15859 .42 .01|1.25 9.90|1.29 9.90|G .41 .52| 45.2 49.1| SME2 |

| 43 31706 15609 -.37 .01|1.26 9.90|1.26 9.90|H .41 .48| 50.7 53.9| RSK1 |

| 49 7112 4999 .81 .02|1.23 9.90|1.25 9.90|I .48 .53| 42.6 46.9| RSK7 |

| 4 32773 15837 -.45 .01|1.20 9.90|1.16 9.90|J .48 .48| 50.2 54.3| SAE4 |

| 46 18192 15439 1.23 .01|1.18 9.90|1.18 9.90|K .51 .55| 43.6 46.3| RSK4 |

| 60 4690 3231 .79 .02|1.13 5.76|1.18 7.41|L .48 .51| 44.6 46.9| RDM9 |

| 53 34798 15801 -.75 .01|1.13 9.90|1.15 9.90|M .40 .46| 53.8 55.3| RDM2 |

| 2 27499 15784 .20 .01|1.12 9.90|1.14 9.90|N .49 .51| 47.9 50.5| SAE2 |

| 30 12719 5015 -1.64 .03|1.14 6.33|1.09 3.35|O .33 .38| 59.3 62.4| SME16|

| 13 7617 5021 .64 .02|1.12 6.20|1.13 7.00|P .48 .52| 45.0 47.7| SAS13|

| 22 30178 15886 -.10 .01|1.09 8.18|1.07 6.10|Q .52 .50| 49.9 52.5| SMG8 |

| 40 7934 5000 .52 .02|1.08 4.47|1.09 4.72|R .52 .51| 49.7 48.6| SOC8 |

| 44 24382 15543 .53 .01|1.09 8.12|1.09 8.26|S .55 .53| 48.1 48.5| RSK2 |

| 8 27963 15654 .12 .01|1.07 6.09|1.05 4.32|T .54 .51| 51.1 51.1| SAE8 |

| 10 28431 15453 .02 .01|1.07 6.37|1.05 4.40|U .50 .50| 54.1 51.8| SAE10|

| 23 28251 15783 .12 .01|1.07 6.81|1.07 6.19|V .54 .51| 48.2 51.2| SMS9 |

| 31 11947 7263 .36 .02|1.07 4.31|1.07 4.46|W .58 .54| 47.7 49.3| SMG17|

| 51 6418 3226 -.23 .03|1.06 2.37|1.07 2.49|X .39 .45| 54.7 54.0| RSK9 |

| 15 21171 15834 .94 .01|1.03 3.23|1.06 5.80|Y .51 .54| 49.3 46.7| SME1 |

| 24 25634 15758 .42 .01|1.04 3.92|1.04 3.87|Z .55 .52| 48.9 49.1| SMS10|

| 47 32437 15502 -.50 .01|1.03 2.32|1.01 1.07| .47 .48| 57.4 54.4| RSK5 |

| BETTER FITTING NOT SHOWN +----------+----------+ | | |

| 37 33093 15588 -.57 .01| .99 -1.29| .97 -2.41| .42 .47| 58.5 54.7| SOC5 |

| 39 11627 5053 -.97 .02| .97 -1.62| .94 -2.68| .41 .43| 61.0 56.2| SOC7 |

| 56 28170 15690 .11 .01| .97 -2.98| .96 -3.49| .51 .51| 53.5 51.2| RDM5 |

| 7 31321 15714 -.29 .01| .96 -3.54| .94 -5.14| .50 .49| 57.4 53.5| SAE7 |

| 45 35186 15598 -.88 .01| .96 -3.21| .93 -5.70| .47 .45| 60.4 55.9| RSK3 |

| 19 26675 15866 .32 .01| .95 -5.32| .95 -4.80|z .55 .52| 51.7 49.7| SMS5 |

| 9 28873 15514 -.02 .01| .94 -5.96| .92 -6.92|y .52 .50| 56.9 52.0| SAS9 |

| 52 27333 15593 .19 .01| .89 -9.90| .94 -5.37|x .43 .51| 55.8 50.6| RDM1 |

| 32 4928 3243 .66 .02| .91 -4.15| .92 -3.47|w .57 .50| 51.0 47.5| SMG18|

| 35 31128 15805 -.24 .01| .90 -8.98| .92 -7.32|v .46 .49| 59.0 53.3| SOC3 |

| 20 27373 15844 .23 .01| .90 -9.90| .90 -9.70|u .55 .51| 54.1 50.3| SMS6 |

| 11 6139 3211 -.07 .02| .88 -5.00| .87 -5.46|t .53 .46| 59.8 53.1| SAE11|

| 17 24304 15890 .60 .01| .86 -9.90| .88 -9.90|s .59 .53| 53.5 48.2| SMS3 |

| 28 24989 15493 .45 .01| .87 -9.90| .87 -9.90|r .58 .52| 54.5 48.9| SMG14|

| 34 29547 15808 -.04 .01| .87 -9.90| .87 -9.90|q .53 .50| 58.1 52.2| SOC2 |

| 57 13795 7248 -.14 .02| .87 -8.45| .85 -9.30|p .58 .52| 58.1 52.2| RDM6 |

| 18 21937 15968 .88 .01| .82 -9.90| .85 -9.90|o .55 .54| 55.0 46.9| SMS4 |

| 58 8938 5022 .16 .02| .85 -8.12| .85 -7.86|n .56 .50| 56.8 51.0| RDM7 |

| 5 32546 15745 -.45 .01| .84 -9.90| .83 -9.90|m .51 .48| 61.7 54.2| SAS5 |

| 59 9467 4984 -.07 .02| .82 -9.89| .80 -9.90|l .59 .49| 60.0 52.7| RDM8 |

| 41 6264 3224 -.13 .03| .81 -8.49| .81 -7.91|k .49 .46| 60.8 53.5| SOC9 |

| 54 29001 15799 .03 .01| .78 -9.90| .79 -9.90|j .53 .50| 60.7 51.7| RDM3 |

| 1 30583 15836 -.16 .01| .75 -9.90| .78 -9.90|i .49 .49| 62.4 52.9| SAS1 |

| 21 27503 15756 .20 .01| .78 -9.90| .78 -9.90|h .57 .51| 58.1 50.5| SMG7 |

| 26 22600 15721 .76 .01| .78 -9.90| .78 -9.90|g .62 .53| 55.3 47.3| SME12|

| 42 7008 3237 -.61 .03| .78 -9.36| .77 -9.16|f .45 .42| 64.9 55.3| SOC10|

| 27 24848 15562 .48 .01| .77 -9.90| .77 -9.90|e .62 .52| 57.7 48.8| SME13|

| 25 25420 15765 .45 .01| .74 -9.90| .75 -9.90|d .61 .52| 58.4 48.9| SMG11|

| 36 31474 15492 -.37 .01| .74 -9.90| .73 -9.90|c .54 .48| 65.7 53.9| SOC4 |

| 6 25451 15714 .44 .01| .70 -9.90| .71 -9.90|b .61 .52| 59.8 49.0| SAS6 |

| 61 6725 3216 -.44 .03| .58 -9.90| .57 -9.90|a .52 .44| 71.8 54.9| RDM10|

|------------------------------------+----------+----------+-----------+-----------+------|

| MEAN 22094.2 12087 .00 .01|1.00 -.7|1.00 -.8| | 54.3 51.8| |

| P.SD 10037.9 5262.3 .55 .01| .19 8.0| .19 7.9| | 6.0 3.1| |

-------------------------------------------------------------------------------------------

Results from a principal component analyses of the residuals are shown in Figure 1.

Figure 1: Principal Components Residual Analyses

------------------------------------------------------------------------------------

STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units

Eigenvalue Observed Expected

Total raw variance in observations = 92.6270 100.0% 100.0%

Raw variance explained by measures = 31.6270 34.1% 33.9%

Raw variance explained by persons = 22.4917 24.3% 24.1%

Raw Variance explained by items = 9.1353 9.9% 9.8%

Raw unexplained variance (total) = 61.0000 65.9% 100.0% 66.1%

Unexplned variance in 1st contrast = 3.3438 3.6% 5.5%

STANDARDIZED RESIDUAL CONTRAST 1 PLOT

-5 -2 1 4 7 10

-+-------------+-------------+-------------+-------------+-------------+- COUNT CLUSTER

| | BA | 2 1

C .5 + |DC + 2 1

O | | |

N .4 + | EF + 2 1

T | |G H | 2 1

R .3 + | IJ + 2 1

A | | L K | 2 1

S .2 + | +

T | ON M | 3 2

.1 + P R + 1 2

1 | | STQ | 4 2

.0 +------------------WV---UZ-X--Y-----------------------------------------+ 6 2

L | 1211 | 5 2

O -.1 + y 1wzx3 + 8 2

A | sqtl2u | 6 3

D -.2 + gmokpn + 6 3

I | jfeih | 6 3

N -.3 + dc | + 2 3

G | ba | 2 3

-.4 + | +

-+-------------+-------------+-------------+-------------+-------------+-

-5 -2 1 4 7 10

ITEM MEASURE

COUNT: 11

1 133671076411

11111111

11246028677007442111

11135067261977387349059429643222 1 13

PERSON 6 11 21325116267717721548560488835518353557748 22 47

T S M S T

%TILE 0 10 30 60 80 90 99

Approximate relationships between the PERSON measures

PCA ITEM Pearson Disattenuated Pearson+Extr Disattenuated+Extr

Contrast Clusters Correlation Correlation Correlation Correlation Cluster Sizes

1 1 - 3 0.6181 0.7247 0.6425 0.7524 12 24

1 1 - 2 0.7273 0.8437 0.7447 0.8625 12 25

1 2 - 3 0.7412 0.8601 0.7644 0.8852 25 24

----------------------------------------------------------------------------------------

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 5: Unidimensionality data for each social and emotional dimension** | | | | |
| Construct | Number of Items | Percent Variance | 1st Contrast Eigenvalue | Cluster Correlations |
| Social and Emotional Learning | 61 | 34.1 | 3.3 | ≥ 0.70 |
| Self-Awareness | 14 | 36.7 | 1.6 | ≥ 0.80 |
| Self-Management | 18 | 41.1 | 2.1 | ≥ 0.80 |
| Social-Awareness | 10 | 35.5 | 1.4 | ≥ 0.80 |
| Relationship Skills | 9 | 49.7 | 1.7 | ≥ 0.70 |
| Responsible Decision-making | 10 | 38.7 | 1.3 | ≥ 0.90 |

Student-level Pearson correlations were evaluated between sub-scale scores for the five separately calibrated dimensions of social and emotional learning. The disattenuated correlations should be positive and of sufficient magnitude (greater than 0.5 but less than 0.9) to indicate that the five sub-scales are measuring distinct but related dimensions of the SE construct. Dimension subscale correlations ranged from 0.62 (between self-management skills and relationship skills) to 0.74 (between self-awareness skills and self-management skills). The results are shown in Table 6 (below the diagonal).

Table 6: Student-level correlations between dimension scores

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Construct\* | SE | SA | SM | SOC | RSK | RDM |
| Social and Emotional Learning (SE) | ----- | 1.00 | 1.00 | 0.97 | 0.94 | 1.00 |
| Self-Awareness (SA) | 0.88 | ----- | 0.86 | 0.84 | 0.85 | 0.90 |
| Self-Management (SM) | 0.92 | 0.74 | ----- | 0.83 | 0.77 | 0.86 |
| Social-Awareness (SOC) | 0.79 | 0.64 | 0.66 | ----- | 0.90 | 0.90 |
| Relationship Skills (RSK) | 0.78 | 0.66 | 0.62 | 0.65 | ----- | 0.95 |
| Responsible Decision-making (RDM) | 0.82 | 0.69 | 0.69 | 0.64 | 0.69 | ----- |

\*Attenuated correlations below diagonal; disattenuated correlations above the diagonal

The evidence from the dimensionality analyses, residual analyses, and the sub-scale correlational analyses supports the structural validity aspect of the social and emotional construct (61 items).

Substantive Validity

Substantive validity assesses whether the responses to the items are consistent with the theoretical framework used to develop the items. Two sets of analyses are used to support the substantive validity aspect of construct validity: these are 1) an examination of the rating scale use by respondents, and 2) an assessment of whether the item difficulty hierarchy of the SELIS survey meets survey developers’ *a priori* expectations.

Rating scale structure. For each threshold of the rating scale, the mean square error fit statistics should be between 0.7 and 1.3. For surveys that use a four-point scale, the distance between the Andrich thresholds should be at least 0.8 logits (Wolfe & Smith, 2007b). In addition, the observed average for each response category should monotonically increase. The rating scale structure is shown in Figure 2 and Table 7.

Figure 2: Rating Scale Structure

------------------------------------------------------------------------------------

CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections

P -+-----------+-----------+-----------+-----------+-----------+-

R 1.0 + 3333333333+

O | 3333333 |

B | 3333 |

A |0 33 |

B .8 + 0 33 +

I | 00 3 |

L | 0 33 |

I | 0 3 |

T .6 + 0 3 +

Y | 0 2222222 3 |

.5 + 0 22 2 3 +

O | 011111 2 \*2 |

F .4 + 110 1\*1 3 2 +

| 1 0 2 1 3 22 |

R | 11 0 2 11 33 2 |

E | 11 2\* 1 3 22 |

S .2 + 11 2 00 \*1 22 +

P |1 22 0 33 1 22 |

O | 22 0\*3 111 2222 |

N | 2222 3333 0000 1111 2222222 |

S .0 +\*\*\*33333333333 000000000\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*+

E -+-----------+-----------+-----------+-----------+-----------+-

-3 -1 1 3 5 7

PERSON [MINUS] ITEM MEASURE

------------------------------------------------------------------------------------

Table 7: Rating scale category information

------------------------------------------------------------------------------------

SUMMARY OF CATEGORY STRUCTURE. Model="R"

---------------------------------------------------------------------

|CATEGORY OBSERVED|OBSVD SAMPLE|INFIT OUTFIT|| ANDRICH |CATEGORY|

|LABEL SCORE COUNT %|AVRGE EXPECT| MNSQ MNSQ||THRESHOLD| MEASURE|

|---------------------+------------+------------++---------+--------|

| 0 0 60023 8| -.61 -.60| 1.00 1.05|| NONE |( -2.67)| 0

| 1 1 175778 24| -.03 .01| .93 .94|| -1.37 | -.92 | 1

| 2 2 332479 45| .66 .62| .88 .89|| -.33 | .79 | 2

| 3 3 169003 23| 1.43 1.47| 1.09 1.08|| 1.70 |( 2.89)| 3

|---------------------+------------+------------++---------+--------|

| MISSING 241096 25| .64 | || | |

------------------------------------------------------------------------------------

The rating scale for the 61 items of the SELIS survey functioned well. The category threshold fit statistics are excellent with MNSQ error near or equal to 1.00 (Figure 2). Adjacent Andrich category thresholds are greater than 0.8 logits apart and the observed average of each response category increases monotonically. Students are using the rating scale structure as intended.

Item hierarchy. The overall item hierarchy across the social and emotional learning scale met DESE’s *a priori* expectations in terms of relative difficulty of individual items across and within dimensions. The ordered pattern of item difficulties for the most part conforms to best test design principles (Wright and Stone, 1979). Figure 3 displays the item-variable map for the SELIS survey with self-awareness (SA) items shown in yellow, self-management (SM) items in orange, social awareness (SOC) items in pink, relationship skills (RSK) in blue, and responsible decision-making (RDM) items shown in green. Items for each dimension span the breadth of the SE continuum with items from different dimensions overlapping as you move from low to high on the scale metric. However, future survey development should examine whether the social awareness scale could be improved by adding “harder” skills and the self-management scale could be improved by adding “easier” skills.

Figure 4 shows the item threshold-variable map; threshold calibrations cover approximately 98.7% of the student distribution. Some gaps in item calibrations are evident toward the top of the student distribution and at the bottom of the distribution. As a result, students at the tail ends of the distribution are measured with more error and are associated with larger standard errors. In terms of average dimension difficulty, students found awareness skills (SA and SOC) relatively easy when compared to their ability to self-manage (SM), form positive and supportive relationships (RSK), and make responsible decisions (RDM) (see Figure 3). This ordering of dimension skills on the social and emotional continuum makes intuitive sense; awareness skills are foundational to being able to self-manage, build positive relationships, and make responsible decisions. Overall, the SELIS survey instrument is well-targeted for the student population in Massashusetts with student and item and item threshold distributions overlapping to a high degree from low to high on the SE scale metric.

Detailed item hierarchies for each of the five dimensions are shown in Figures 5 (SA), 6 (SM), 7 (SOC), 8 (RSK), and 9 (RDM). Each item hierarchy met DESE’s *a priori* expectations and with one item exception, conformed to those of the developers of the SECA instrument (Davidson et al., 2018; Crowder et al., 2019) upon which SELIS is founded. The social awareness item, “Knowing how my actions impact my classmates” (SOC4), was easier than expected for respondents; the expectation was that it would be more difficult skill than, “Know what people may be feeling by the look on their face” (SOC 1), but this item (SOC 1) was the more difficult skill of the two items for students and the ordering of the social awareness items was inverted for these two items. The remaining social awareness items followed the expected item ordering when compared to the original developers’ findings.

Figure 3: SELIS item-variable map

**MEASURE    PERSON - MAP - ITEM  
                <more>│<rare>  
     6             .  ┼  
                   .  │  
                   .  │  
                   .  │  
     5             .  ┼  
                      │  
                   .  │  
                   .  │  
                   .  │  
     4             .  ┼  
                   .  │  
                   .  │  
                   .  │  
                   .  │  
                   .  │  
     3             .  ┼  
                   .  │  
                   .  │  
                  .# T│  
                  .#  │  
     2            .#  ┼  
                 .##  │  
                .###  │  
               .#### S│  
              .#####  │  RSK4  
            .#######  │T  
     1     .########  ┼  SME1   SMS4  
       .############  │  RDM9   RSK7   SME12  
        .########### M│S RSK2   SAS13  SMG18  SMS3   SOC8  
       .############  │  SAS6   SME13  SME2   SMG11  SMG14  SMG17  SMS10  SMS5  
        .###########  │  RDM1   RDM7   RSK8   SAE2   SMG7   SMS6  
     0  .###########  ┼M RDM3   RDM5   SAE10  SAE8   SAS9   SME15  SMS9   SOC2  
           .########  │  RDM6   RDM8   SAE11  SAS1   SAS14  SMG8   SOC1   SOC9  
             .###### S│  RSK1   RSK9   SAE3   SAE7   SOC3   SOC4  
               .####  │S RDM10  RSK5   SAE4   SAS5   SOC5   SOC6  
                .###  │  RDM2   RDM4   SOC10  
                 .##  │  RSK3   SAS12  
    -1            .#  ┼T SOC7  
                   . T│  RSK6  
                   .  │  
                   .  │  SME16  
                   .  │  
    -2             .  ┼  
                   .  │  
                   .  │  
                   .  │  
                   .  │  
                   .  │  
    -3             .  ┼  
                   .  │  
                   .  │  
                      │  
                      │  
    -4             .  ┼  
                   .  │  
    -5             .  ┼  
                <less>│<freq>  
  EACH "#" IS 120: EACH "." IS 1 TO 119** Note: Map has been truncated (top and bottom) to fit to the page

Figure 4: Item-threshold variable map

   
  MEASURE               │ BOTTOM P=50%  │ MEASURE       │ TOP P=50%    MEASURE  
   <more> ───── PERSON ─┼─ ITEM        ─┼─ ITEM        ─┼─ ITEM         <rare>  
     6                . ┼               ┼               ┼                  6  
                      . │               │               │  
                      . │               │               │  
                        │               │               │  
                        │               │               │  
                      . │               │               │  
     5                . ┼               ┼               ┼                  5  
                        │               │               │  
                      . │               │               │  
                      . │               │               │  
                      . │               │               │  
     4                . ┼               ┼               ┼                  4  
                      . │               │               │  
                      . │               │               │  
                      . │               │               │  
                      . │               │               │  
                      . │               │               │ X  
     3                . ┼               ┼               ┼                  3  
                      . │               │               │ XX  
                      . │               │               │ XXX  
                     .# │               │               │ XXXXX  
                     .# │               │               │ XXXXXXXX  
     2               .# ┼               ┼               ┼ XXXXXX           2  
                    .## │               │               │ XXXXXXXX  
                   .### │               │               │ XXXXXXXX  
                  .#### │               │               │ XXXXXX  
                 .##### │               │ X             │ XXXXXX  
               .####### │               │               │ XXX  
     1        .######## ┼               ┼ XX            ┼ XX               1  
          .############ │               │ XXX           │ X  
           .########### │               │ XXXXX         │ X  
          .############ │               │ XXXXXXXX      │  
           .########### │               │ XXXXXX        │ X  
     0     .########### ┼               ┼ XXXXXXXX      ┼                  0  
              .######## │               │ XXXXXXXX      │  
                .###### │ X             │ XXXXXX        │  
                  .#### │               │ XXXXXX        │  
                   .### │ XX            │ XXX           │  
                    .## │ XXX           │ XX            │  
    -1               .# ┼ XXXXX         ┼ X             ┼                 -1  
                      . │ XXXXXXXX      │ X             │  
                      . │ XXXXXX        │               │  
                      . │ XXXXXXXX      │ X             │  
                      . │ XXXXXXXX      │               │  
    -2                . ┼ XXXXXX        ┼               ┼                 -2  
                      . │ XXXXXX        │               │  
                      . │ XXX           │               │  
                      . │ XX            │               │  
                      . │ X             │               │  
                      . │ X             │               │  
    -3                . ┼               ┼               ┼                 -3  
                      . │ X             │               │  
                      . │               │               │  
                        │               │               │  
                        │               │               │  
    -4                . ┼               ┼               ┼                 -4  
                        │               │               │  
                      . │               │               │  
                        │               │               │  
                        │               │               │  
                        │               │               │  
    -5                . ┼               ┼               ┼                 -5  
   <less> ───── PERSON ─┼─ ITEM        ─┼─ ITEM        ─┼─ ITEM         <freq>  
  EACH "#" IN THE PERSON COLUMN IS 120 PERSONS: EACH "." IS 1 TO 119

HARDEST SKILLS TO DO EASIEST SKILLS TO DO

|  |  |  |
| --- | --- | --- |
| **Item** | **Domain** | **Item Prompt (from most difficult to least difficult)** |
| SAS13 | Self-concept | Offering an answer even though I am not sure I am right. |
| SAS6 | Self-concept | Knowing how to get better at things that are hard for me to do at school. |
| SAE2 | Emotions | Knowing ways to make myself feel better when I'm sad. |
| SAE8 | Emotions | Knowing ways I calm myself down. |
| SAE10 | Emotions | Knowing when my feelings are making it hard for me to focus. |
| SAS9 | Self-concept | Knowing when I can’t control something. |
| SAE11 | Emotions | Understanding the difference between how others expect me to feel and how I really feel. |
| SAS14 | Self-concept | Knowing my opinions matter even when others don't listen to them. |
| SAS1 | Self-concept | Knowing what my strengths are. |
| SAE7 | Emotions | Knowing when my mood affects how I treat others. |
| SAE3 | Emotions | Noticing what my body does when I am nervous. |
| SAE4 | Emotions | Knowing the emotions I feel. |
| SAS5 | Self-concept | Knowing when I am wrong in the way I behave. |
| SAS12 | Self-concept | Knowing what I am interested in at school. |

HARDEST SKILLS TO DO EASIEST SKILLS TO DO

|  |  |  |
| --- | --- | --- |
| **Item** | **Domain** | **Item Prompt (from most difficult to least difficult)** |
| SME1 | Emotion | Staying calm when I feel stressed. |
| SMS4 | Schoolwork | Staying focused in class even when there are distractions. |
| SME12 | Emotion | Getting through something even when I feel frustrated. |
| SMG18 | Goal-setting | Planning my work when I have multiple assignments due at the same time. |
| SMS3 | Schoolwork | Working on assignments even when they are hard. |
| SME13 | Emotion | Working on things even when I don't like them. |
| SMG11 | Goal-setting | Finishing tasks even if they are hard for me. |
| SMG14 | Goal-setting | Reaching goals that I set for myself. |
| SMS10 | Schoolwork | Doing my schoolwork even when I do not feel like it. |
| SME2 | Emotion | Being patient even when I am really excited. |
| SMG17 | Goal-setting | Catching up on my work when I get behind. |
| SMS5 | Schoolwork | Planning ahead so I can turn a project in on time. |
| SMS6 | Schoolwork | Being prepared for tests. |
| SMG7 | Goal-setting | Thinking through the steps it will take to reach my goals. |
| SMS9 | Schoolwork | Finishing my schoolwork without reminders. |
| SMG8 | Goal-setting | Setting goals for myself. |
| SME16 | Emotion | Being careful about what information I share about myself on social media (e.g., TikTok, Facebook, Instagram, etc.). |

HARDEST SKILL TO DO EASIEST SKILL TO DO

|  |  |
| --- | --- |
| **Item** | **Item Prompt (from most difficult to least difficult)** |
| SOC8 | Knowing when to tell my friends how I really feel. |
| SOC2 | Knowing how to get help when I'm having trouble with a classmate. |
| SOC9 | Knowing when my tone of voice may cause someone to misunderstand what I intended to say. |
| SOC1 | Knowing what people are feeling by the look on their face. |
| SOC3 | Learning from people with different opinions than me. |
| SOC4 | Knowing how my actions impact my classmates. |
| SOC6 | Waiting for other students to finish talking before I speak. |
| SOC5 | Knowing when someone needs help. |
| SOC10 | Recognizing when I am making someone uncomfortable. |
| SOC7 | Knowing the difference between bullying someone and joking with someone. |

HARDEST SKILL TO DO EASIEST SKILL TO DO

|  |  |
| --- | --- |
| **Item** | **Item Prompt (from most difficult to least difficult)** |
| RSK4 | Sharing what I am feeling with others. |
| RSK7 | Leading a discussion in class. |
| RSK2 | Talking to an adult when I have problems at school. |
| RSK8 | Asking a classmate to do their fair share of a group project. |
| RSK9 | Standing up for someone even if they are outside of my friend group. |
| RSK1 | Being welcoming to someone I don't usually eat lunch with. |
| RSK5 | Respecting a classmate's opinions during a disagreement. |
| RSK3 | Getting along well with my classmates. |
| RSK6 | Getting along well with my teachers. |

HARDEST SKILL TO DO EASIEST SKILL TO DO

|  |  |
| --- | --- |
| **Item** | **Item Prompt (from most difficult to least difficult)** |
| RDM9 | Deciding what courses to take to get me into the college or career I want. |
| RDM1 | Helping to make my school a better place. |
| RDM7 | Knowing when to accept help when it is offered. |
| RDM5 | Thinking about what might happen before making a decision. |
| REM3 | Thinking of different ways to solve a problem (e.g., a disagreement with another student). |
| RDM8 | Understanding the importance of what I am learning to my future success. |
| RDM6 | Taking responsibility for my choices. |
| RDM10 | Knowing something is wrong to do even when it happens all the time. |
| RDM4 | Saying "no" to a friend who wants to break the rules. |
| RDM2 | Knowing when people's actions are "right" or "wrong" (e.g., helpful or harmful) |

GENERALIZABILITY: All students (N = 16,039)

A measure is considered generalizable when the score meaning and properties function similarly across multiple contexts (e.g., districts, student groups). Reliability analyses and differential item functioning (DIF) analyses are used to assess the generalizability of the measures. Similar to Cronbach’s alpha, person separation reliability (PSR) looks at the stability (internal consistency) of the measures across scoring structures. DESE used DIF analyses to empirically test for item invariance across districts and several student groups; item invariance ensures comparability of score interpretation.

Reliability analyses. Best test design principles (Wright, 1979) necessitate the alignment of the mean of the item distribution to the mean of the person distribution. The mean person difficulty of the 61-item scale was +0.60 logits with a standard deviation of 1.00 logits (Table 8). The items are reasonably well targeted for the student distribution (Table 8) resulting in a real person separation reliability (PSR) of 0.93, and a real person separation index of 3.63. Only 53 students (0.3%) had a maximum extreme score with less than 0.1% of students (5 students) with a minimum extreme score.

Tables 9 and 10 summarize the reliability data by SE dimension and by student group, respectively. Dimensions composed of fewer items (social awareness, relationship skills, and responsible decision-making) have, on average, lower reliabilities, albeit all over 0.7. These lower reliabilities mean that students can only be statistically and reliably divided into two scoring groups or person strata (high and low). In contrast, the higher reliabilities for the self-awareness and self-management dimensions can be reliably divided into 3 and 4 scoring groups, respectively. The reliabilities at the student group level for the SE construct as a whole and for each dimension are, for the most part, sufficient to accurately assess group difference; reliabilities lower than 0.7 are highlighted in brown (Table 10).

Table 8: Reliability data for SE persons (students) and items

-------------------------------------------------------------------------------

SUMMARY OF 15980 MEASURED (NON-EXTREME) PERSON

-------------------------------------------------------------------------------

| TOTAL MODEL INFIT OUTFIT |

| SCORE COUNT MEASURE S.E. MNSQ ZSTD MNSQ ZSTD |

|-----------------------------------------------------------------------------|

| MEAN 83.9 46.0 .58 .21 1.01 -.37 1.01 -.36 |

| SEM .2 .0 .01 .00 .00 .02 .00 .02 |

| P.SD 21.6 4.2 .92 .04 .54 2.66 .54 2.64 |

| S.SD 21.6 4.2 .92 .04 .54 2.66 .54 2.64 |

| MAX. 149.0 50.0 5.74 1.01 4.03 9.35 9.90 9.38 |

| MIN. 3.0 20.0 -4.20 .18 .08 -8.85 .08 -8.55 |

|-----------------------------------------------------------------------------|

| REAL RMSE .24 TRUE SD .89 SEPARATION 3.68 PERSON RELIABILITY .93 |

|MODEL RMSE .22 TRUE SD .89 SEPARATION 4.08 PERSON RELIABILITY .94 |

| S.E. OF PERSON MEAN = .01 |

-------------------------------------------------------------------------------

MAXIMUM EXTREME SCORE: 53 PERSON .3%

MINIMUM EXTREME SCORE: 6 PERSON .0%

SUMMARY OF 16039 MEASURED (EXTREME AND NON-EXTREME) PERSON

-------------------------------------------------------------------------------

| TOTAL MODEL INFIT OUTFIT |

| SCORE COUNT MEASURE S.E. MNSQ ZSTD MNSQ ZSTD |

|-----------------------------------------------------------------------------|

| MEAN 84.0 46.0 .60 .22 1.01 -.4 1.01 -.4 |

| SEM .2 .0 .01 .00 |

| P.SD 21.9 4.2 1.00 .11 |

| S.SD 21.9 4.2 1.00 .11 |

| MAX. 150.0 50.0 6.96 1.83 |

| MIN. .0 20.0 -6.59 .18 |

|-----------------------------------------------------------------------------|

| REAL RMSE .26 TRUE SD .96 SEPARATION 3.63 PERSON RELIABILITY .93 |

|MODEL RMSE .24 TRUE SD .97 SEPARATION 3.94 PERSON RELIABILITY .94 |

| S.E. OF PERSON MEAN = .01 |

-------------------------------------------------------------------------------

PERSON RAW SCORE-TO-MEASURE CORRELATION = .90

CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .93 SEM = 5.89

STANDARDIZED (50 ITEM) RELIABILITY = .93

SUMMARY OF 61 MEASURED (NON-EXTREME) ITEM

-------------------------------------------------------------------------------

| TOTAL MODEL INFIT OUTFIT |

| SCORE COUNT MEASURE S.E. MNSQ ZSTD MNSQ ZSTD |

|-----------------------------------------------------------------------------|

| MEAN 22094.2 12086.6 .00 .01 1.00 -.71 1.00 -.77 |

| SEM 1295.9 679.4 .07 .00 .02 1.03 .02 1.02 |

| P.SD 10037.9 5262.3 .55 .01 .19 7.97 .19 7.87 |

| S.SD 10121.2 5306.0 .56 .01 .19 8.04 .19 7.93 |

| MAX. 37477.0 15968.0 1.23 .03 1.41 9.90 1.40 9.90 |

| MIN. 4690.0 3211.0 -1.64 .01 .58 -9.90 .57 -9.90 |

|-----------------------------------------------------------------------------|

| REAL RMSE .02 TRUE SD .55 SEPARATION 34.05 ITEM RELIABILITY 1.00 |

|MODEL RMSE .02 TRUE SD .55 SEPARATION 35.18 ITEM RELIABILITY 1.00 |

| S.E. OF ITEM MEAN = .07 |

-------------------------------------------------------------------------------

Table 9: SEL construct and dimension reliabilities

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Construct/Dimension | Number of Items | Real  PSR1 | Model  PSR1 | PSI  average2 | Number of Person Strata3 |
| Social and Emotional Learning (SE) | 61 | 0.93 | 0.94 | 3.80 | 5.4 |
| Self-Awareness (SA) | 14 | 0.79 | 0.82 | 2.0 | 3.0 |
| Self-Management (SM) | 18 | 0.87 | 0.89 | 2.75 | 4.0 |
| Social-Awareness (S0C) | 10 | 0.63 | 0.71 | 1.43 | 2.2 |
| Relationship Skills (RSK) | 9 | 0.67 | 0.73 | 1.54 | 2.4 |
| Responsible Decision-making (RDM) | 10 | 0.65 | 0.72 | 1.49 | 2.3 |

1PSR: Person Separation Reliability; 2PSI: Person Separation Index; 3Formula: (((PSI \* 4)+1)/3)

Table 10: Construct and dimension reliabilities by student group

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Student Group  (Number) | Construct/Dimension  Person Separation Reliability (Model) | | | | | |
| SE | SA | SM | SOC | RSK | RDM |
| Female (8082) | 0.94 | 0.83 | 0.90 | 0.70 | 0.74 | 0.72 |
| Male (7949) | 0.94 | 0.81 | 0.89 | 0.71 | 0.72 | 0.72 |
| Non-binary (NA) | NA | NA | NA | NA | NA | NA |
| Asian (762) | 0.93 | 0.80 | 0.87 | 0.67 | 0.70 | 0.68 |
| Black (948) | 0.94 | 0.82 | 0.89 | 0.71 | 0.73 | 0.70 |
| Hispanic (3642) | 0.94 | 0.83 | 0.89 | 0.72 | 0.74 | 0.73 |
| Other (718) | 0.94 | 0.83 | 0.89 | 0.71 | 0.72 | 0.71 |
| White (9968) | 0.94 | 0.82 | 0.89 | 0.70 | 0.73 | 0.72 |
| Non-Economically disadvantaged (9979) | 0.94 | 0.81 | 0.89 | 0.69 | 0.72 | 0.70 |
| Economically disadvantaged (6059) | 0.94 | 0.83 | 0.90 | 0.72 | 0.74 | 0.74 |
| Non-English Learner (14882) | 0.94 | 0.82 | 0.89 | 0.70 | 0.73 | 0.71 |
| English Learner (1156) | 0.94 | 0.83 | 0.88 | 0.73 | 0.73 | 0.75 |
| Students without disabilities (13408) | 0.94 | 0.82 | 0.89 | 0.70 | 0.73 | 0.71 |
| Students with disabilities (2630) | 0.94 | 0.84 | 0.90 | 0.74 | 0.74 | 0.75 |

Differential Item Function (DIF) Analyses. To support the claim that the SELIS survey data is generalizable, the items should have the same meaning (as measured by item difficulty) for different groups of respondents (e.g., gender, race/ethnicity). Respondents of the same SE skill level, should have the same probability of affirming an item irrespective of the group they belong to. In this study, items were flagged if the average difficulties differed by 0.5 logits or more. DIF plots for gender (Figure 10), race/ethnicity (Figures 11 and 12), economically disadvantaged status (Figure 13), English learner status (Figure 14), and students with disabilities status (Figure 15) are shown. Each figure is accompanied by a table indicating the number of items within each item difficulty measure difference category.

Items were largely invariant across each student group analysis. The analyses indicated that item deltas (difficulties) did not differ significantly across the following student groups: gender, race/ethnicity (when white students were compared to all students of color), economically disadvantaged status, and students with disabilities status. One item (RDM2) exhibited moderate DIF (difference equals 0.5 to 0.6 logits) when comparing non-English learners to English learners (EL); EL students found this skill harder when compared to non-EL students of the same ability level. When race/ethnicity data were broken out into five student groups, one item (RSK9) exhibited moderate DIF with an additional item exhibiting severe DIF (SAS14, ≥0.7 logits). In both items, Asian students found the skills harder when compared to their peers of similar ability. DIF items should be reviewed for revision. If kept as is, they can be removed from the calibration process when producing scores for these student groups; these three items function well across most student group comparisons and are largely productive for measurement.

The item difficulties should also be largely invariant across the 10 districts when each districts’ student responses are calibrated separately. A matrix plot (Figure 16) shows the Pearson correlations of the 61 items across the 10 districts. The correlations ranged from 0.93 to 0.99 indicating a strong relationship between the item difficulty calibrations. The weaker correlations were between districts that were smaller in size; the precision of item estimates was lower for these districts.

The reliability data for the overall SELIS scale and the reliability data for the five dimensions support the generalizability of the construct and associated dimensions. A large majority of items exhibited no DIF across five different student group comparisons and across 10 district comparisons providing further evidence for the generalizability of the scores.

Figure 10: DIF plot for gender comparison1

1Insufficient number of non-binary students to include in analyses

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Size of DIF (difference in logits) | | | | | |
|  | <0.3 | 0.3 to <0.4 | 0.4 to <0.5 | 0.5 to <0.6 | 0.6 to <0.7 | ≥0.7 |
| Number of items | 52 | 5 | 4 | 0 | 0 | 0 |
| Items | ----- | ----- | ---- | ----- | ----- | ----- |

Figure 11: DIF plot by race/ethnicity (White versus Students of Color) comparison

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Size of DIF (difference in logits) | | | | | |
|  | <0.3 | 0.3 to <0.4 | 0.4 to <0.5 | 0.5 to <0.6 | 0.6 to <0.7 | ≥0.7 |
| Number of items | 61 | 0 | 0 | 0 | 0 | 0 |
| Items | ------ | ------ | ------ | ------ | ------ | ------ |

Figure 12: DIF plot for race/ethnicity (5 student groups) comparison

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Size of DIF (difference in logits) | | | | | |
|  | <0.3 | 0.3 to <0.4 | 0.4 to <0.5 | 0.5 to <0.6 | 0.6 to <0.7 | ≥0.7 |
| Number of items | 47 | 7 | 3 | 2 | 1 | 1 |
| Items | ------ | ------ | ------ | ------ | RSK9 | SAS14 |
| Comment | ------ | ------ | ------ | ------ | Asian students view skill more difficult | Asian students view skill more difficult |
| DIF Items | RSK9: Standing up for someone even if they are outside of my friend group.  SAS14: Knowing my opinions matter even when others don't listen to them. | | | | | |

Figure 13: DIF plot for economically disadvantaged status comparison

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Size of DIF (difference in logits) | | | | | |
|  | <0.3 | 0.3 to <0.4 | 0.4 to <0.5 | 0.5 to <0.6 | 0.6 to <0.7 | ≥0.7 |
| Number of items | 60 | 0 | 1 | 0 | 0 | 0 |
| Items | ------ | ------ | ------ | ------ | ------ | ------ |

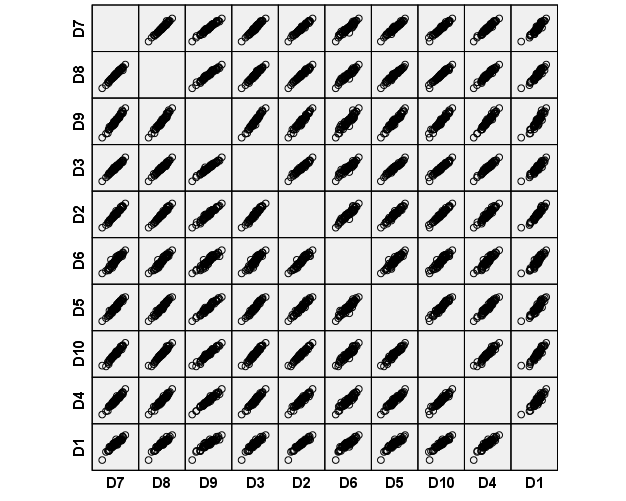
Figure14: DIF plot for English learner status comparison

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Size of DIF (difference in logits) | | | | | |
|  | <0.3 | 0.3 to <0.4 | 0.4 to <0.5 | 0.5 to <0.6 | 0.6 to <0.7 | ≥0.7 |
| Number of items | 57 | 2 | 1 | 0 | 1 | 0 |
| Items | ------ | ------ | ----- | ------ | RDM2 | ------ |
| Comment |  |  |  |  | EL students view the skill harder |  |
| DIF item | RDM2: Knowing when people's actions are "right" or "wrong" (e.g., helpful, or harmful). | | | | | |

Figure 15: DIF plot for students with disabilities status comparison

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Size of DIF (difference in logits) | | | | | |
|  | <0.3 | 0.3 to <0.4 | 0.4 to <0.5 | 0.5 to <0.6 | 0.6 to <0.7 | >0.7 |
| Number of items | 58 | 1 | 2 | 0 | 0 | 0 |
| Items | ------ | ------ | ------ | ------ | ----- | ------ |

Figure 16: Matrix plot of district-level item difficulty correlations (61 items)



External Validity

This aspect of construct validity relates to the responsiveness of an instrument and the relationship of its scores to the scores of external measures (criterion validity). The responsiveness of an instrument refers to “the degree to which an instrument is capable of detecting changes in person measures following an intervention that is assumed to impact the target construct” (Wolfe & Smith, 2007b, p. 222). If an instrument is responsive, it can be applied appropriately to measure expected group differences or individual/group change. Responsiveness is examined at the student-level and school level.

Criterion validity is the strongest form of external validity; it determines how well scores from an instrument predict scores on a criterion measure (e.g., how well do SELIS scores predict achievement or the reverse, how well do prior achievement scores predict SELIS scores). There are two forms of criterion validity: concurrent and predictive. This section reports preliminary data on the relationship between students’ achievement (2019 MCAS scores) and students’ 2021 SELIS scores. The research question asks: Does student prior achievement (2019 MCAS scores) predict students’ perceptions of their social and emotional learning skills (SELIS 2021 scores)? Unfortunately, at the time of this posting, students 2021 achievement scores were not available so concurrent validity evidence could not be ascertained.

Responsiveness of SELIS instrument.

Student-level. The responsiveness of an instrument is measured by the person strata index, H, which provides the number of statistically distinct SE skill groups whose centers of score distributions are separated by at least three standard errors of measurement within the sample. The number of person strata for the 61-item SELIS instrument is equivalent to almost 5.4 distinct person strata indicating that the student score distribution can reliably be divided into 5 distinct levels (Table 11). Students were categorized into four skill levels in the district reports, and this provides evidence that this score division is supported as the SELIS instrument proved responsive. The responsiveness of each dimension was also examined to assess if they could potentially be used as standalone administrations. Dimensions with a higher number of items were most responsive and students can be reliably divided into 3 or 4 skill levels for self-awareness and self-management, respectively. As mentioned previously, the three dimensions’ score distributions that are predicated on a smaller number of items (social awareness, relationship skills, and responsible decision-making skills) can only reliably be divided students into two skill levels (high and low). These data suggest, particularly for the overall SE score distribution, that the SELIS instrument is responsive at the student-level.

Table 11: SE construct and dimension reliabilities

|  |  |  |
| --- | --- | --- |
| Construct/Dimension | Number of Items | Number of Person Strata3 |
| Social and Emotional Learning (SE) | 61 | 5.4 |
| Self-Awareness (SA) | 14 | 3.0 |
| Self-Management (SM) | 18 | 4.0 |
| Social-Awareness (S0C) | 10 | 2.2 |
| Relationship Skills (RSK) | 9 | 2.4 |
| Responsible Decision-making (RDM) | 10 | 2.3 |

1PSR: Person Separation Reliability; 2PSI: Person Separation Index; 3Formula: (((PSI \* 4)+1)/3)

School-level Responsiveness. The greater the number of person strata at the individual-level, the more likely the instrument will be responsive at the school level. The mean of the school-level data was 0.56 logits with a standard deviation of 0.164 logits. The average SE construct score ranged from low of 0.10 logits to a high of 0.97 logits (a 5.4 standard deviation difference). The distribution of school-level scores is shown in Figure 17. Percentiles are provided in Table 12. The 50th percentile is 0.58 logits; the 25th and 75th percentiles are 0.47 and 0.66 logits, respectively. These data suggest that the SELIS instrument is responsive at the school-level.

Figure 17: School-level distribution of SE construct scores

Chart, histogram

Histogram shows the school-level social and emotional construct scaled-score distribution. 82 schools contribute to the distribution.

Table 12: Percentile range of SE construct scores

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Percentiles** | | | | | | | | |
|  | | Percentiles | | | | | | |
| 5 | 10 | 25 | 50 | 75 | 90 | 95 |
| Weighted Average(Definition 1) | DSEms | .2200 | .3160 | .4675 | .5800 | .6650 | .7570 | .8100 |
| Tukey's Hinges | DSEms |  |  | .4700 | .5800 | .6600 |  |  |

Responsiveness at the school-level is important as the scores can be used to help determine if Tier 1 interventions are needed within certain schools within districts.

Predictive Validity Evidence.

These analyses examine if student prior achievement (2019) is predictive of their 2021 social and emotional perception scores. Static achievement was measured using the Massachusetts Comprehensive Assessment System (MCAS) scaled scores from 2019 for English Language Arts & Literacy (ELA) and for mathematics. In separate analyses, student growth scores were also used as predictors; do students’ 2019 ELA and, separately, mathematics growth scores predict their 2021 SELIS scores? All achievement scaled scores and SELIS scores were standardized for these analyses. Bivariate correlations between student achievement were examined first at both the student- and school-level. *Please note that there are only 36 schools retained in the growth analyses. Most elementarys are excluded from the analyses as there are no growth scores available in 2019 for students who are in grades 3, 4, and 5 in 2021.*  Regression analyses to determine if any significant Pearson correlations between student-level achievement and SELIS scores were sustained when student covariates were added to the regression model were also performed.

Pearson Correlations. With the exception of three correlations, the student-level Pearson correlations are small but significant (Table 13) across each bivariate correlation. This indicates that, for the most part, students’ 2019 ELA and mathematics static achievement scores (Zescaleds and Zmscaleds) and growth scores (Zesgp and Zmsgp) are predictive of students’ 2021 SE scores. In terms of core competencies, the strongest correlations are between students’ four achievement scores and students’ self-management scores. The weakest or non-significant correlations are between achievement and the two awareness competencies (SA and SOC).

Table 13: Student-level Pearson Correlations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Spring 2021 SELIS score | 2019 MCAS Zescaleds1  (N = 10,455) | 2019 MCAS Zesgp2  (N = 7,136) | 2019 MCAS  Zmscaleds3  (N = 10,503) | 2019 MCAS  Zmsgp4  (N = 7,138) |
| SE construct  (ZSEms) | 0.095\*\* | 0.033\*\* | 0.106\*\* | 0.039\*\* |
| Self-awareness  (ZSAms) | 0.041\*\* | 0.008 | 0.062\*\* | 0.015 |
| Self-management  (ZSMms) | 0.103\*\* | 0.040\*\* | 0.131\*\* | 0.061\*\* |
| Social awareness  (ZSOCms) | 0.079\*\* | 0.026\* | 0.043\*\* | 0.000 |
| Relationship skills  (ZRSKms) | 0.067\*\* | 0.029\* | 0.062\*\* | 0.029\* |
| Responsible decision-making  (ZRDMms) | 0.102\*\* | 0.035\*\* | 0.090\*\* | 0.026\* |

1escaleds: English Language Arts & Literacy Z scaled score; 2English Language Arts and Literacy student group percentile Z score 3mscaleds: Mathematics Z scaled score; 4Mathematics student growth percentile Z score.

When aggregated to the school level, there are significant small-to-moderate (*r* = 0.30 to 0.50) correlations between students’ static achievement scores and the SE construct/competency scores (Table 14). The analyses related to growth scores were problematic as they only included 36 of the 82 schools in the sample. All of the math growth score correlations were not significant; the ELA growth scores were only associated with the overall SE construct score and the self-management and responsible decision-making competency scores. Hough, Kalogrides, & Loeb (2017) similarly found that, at the school-level, students’ concurrent static mathematics achievement scores were correlated with students’ SE scores across all school types (elementary, middle, and high) but not meaningfully correlated with their mathematics growth scores. There was insufficient data to reliably break the SELIS data out by school type. However, the magnitude of the correlations found in this study correspond to those determined by Hough, Kalogrides, & Loeb (2017).

Table 14: School-level Pearson Correlations1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Spring 2021 SELIS score | 2019 MCAS Zescaleds  (N = 73) | 2019 MCAS Zesgp1  (N = 36) | 2019 MCAS  Zmscaleds  (N = 73) | 2019 MCAS  Zmsgp1  (N = 36) |
| SE construct  (ZSEms) | 0.40\*\* | 0.36\* | 0.43\* | 0.24 |
| Self-awareness  (ZSAms) | 0.35**\*\*** | 0.32 | 0.41\*\* | 0.15 |
| Self-management  (ZSMms) | 0.33\*\* | 0.39\* | 0.36\*\* | 0.27 |
| Social awareness  (ZSOCms) | 0.50\*\* | 0.25 | 0.47\*\* | 0.17 |
| Relationship skills  (ZRSKms) | 0.34\*\* | 0.27 | 0.30\*\* | 0.24 |
| Responsible decision-making  (ZRDMms) | 0.43\*\* | 0.33\* | 0.47\*\* | 0.20 |

1Caution is advised as there are only 36 schools associated with the student growth percentile analyses.

Regression Analyses. Intra-class correlations (ICCs) were calculated for the overall SE scaled score and the five dimension scaled scores (Appendix D). Less than 5% of the variance is explained at the school or teacher level when performing two-level models. Similarly, using a three level model which includes both school (Level 2) and teacher (Level 3) in the model, roughly 1.3% of variance in SELIS scores is explained by school-level factors with approximately 3.0% explained by teacher-level factors. Due to these low ICCs, multi-level modeling was not warranted as over 95 percent of the variation in SELIS scores is at the individual student level. Simple multi-linear regressions were carried out to determine if students’ 2019 achievement scores predict students’ 2021 SELIS scores after partialling out the variance due to student characteristics. Table 15 provides the regression results of models predicting the SE construct/dimensions with student achievement and student covariates. Appendix E provides complete regression results related to the SE construct score (full dimension regression results are not shown).

Table 15: Standardized regression coefficients with student covariates included

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Spring 2021 SELIS score | 2019 MCAS Zescaleds1  (N = 10,450) | 2019 MCAS  Zesgp3  (N = 7,132) | 2019 MCAS Zmscaleds2  (N = 10,498) | 2019 MCAS  Zmsgp4  (N = 7,138) |
| SE construct  (ZSEms) | 0.061\*\*\* | 0.032\*\* | 0.062\*\*\* | 0.032\*\* |
| Self-awareness  (ZSAms) | 0.024\* | 0.014 | 0.025\* | 0.014 |
| Self-management  (ZSMms) | 0.078\*\*\* | 0.039\*\* | 0.100\*\*\* | 0.055\*\*\* |
| Social awareness  (ZSOCms) | 0.042\*\*\* | 0.020 | 0.000 | -0.006 |
| Relationship skills  (ZRSKms) | 0.043\*\*\* | 0.027\* | 0.029\* | 0.023 |
| Responsible decision-making  (ZRDMms) | 0.052\*\*\* | 0.028\* | 0.040\*\* | 0.017 |

Students’ 2019 ELA and mathematics achievement scores (static or growth-related) are still predictive of students’ 2021 social and emotional skills (ZSEms) after partialling out the variance due to student characteristics (Table 15). In order of magnitude, students’ ELA static performance has the strongest relationship with self-management skills, responsible decision-making skills, relationship skills, social awareness skills and a relatively weak relationship with self-awareness skills. The order of magnitude is similar for static mathematics scores with the exception of social awareness skills; math achievement is not predictive of students’ social awareness skills.

The relationship between students’ growth scores and social and emotional skills is weaker (when compared to static scores) but still significant after partialling out the variance due to student characteristics. Similar to static achievement, students’ ELA and mathematics growth scores have the strongest relationship with students’ self-management scores; for mathematics, the relationship between growth and self-management scores is the only significant relationship between students’ math growth and students’ core competency skills. ELA growth scores are not predictive of students’ self-awareness or social awareness skills but are predictive of students’ relationship skills and responsible decision-making skills. These analyses need repeated to determine the concurrent relationship between students’ 2021 achievement scores and their 2021 social and emotional scores (student 2021 achievement data was not available at the time of the writing of this study).

Consequential Validity

Consequential validity discusses the implications of using the scores for their intended purpose. It “appraises the value implications of score interpretation as a basis for action as well as the actual and potential consequences of test use” (Messick, 1995b, p.6). The Massachusetts Safe and Supportive Schools Commission (2019a, p.1) advocates that, “safe and supportive school environments are essential in order to reach high academic standards and other important educational reform goals, including diminishing the use of suspension and expulsion as an approach to discipline, preventing bullying, preventing substance use and misuse and providing support for addiction recovery, closing proficiency gaps, and halting the school to prison pipeline.” DESE’s primary goal is for educators to use the SELIS data to support students’ social and emotional wellbeing within a multi-tiered system of student supports.

These data do have moderate stakes attached. Although they are not used in a high stakes accountability system at the state or district level, districts and schools are tasked with interpreting the data in culturally responsive ways to best support all their students, including students who are often marginalized. If the data is used inappropriately by educators, for example, to confirm previous biases about their students, students could be negatively impacted, and their social and emotional development harmed instead of supported. Involving students in the co-interpretation of the SELIS data and SWON maps is recommended so students’ have a voice in setting their goals for their SE development.

Conclusion and Limitations of Study

The purpose of this validity study was to provide psychometric evidence to justify the use of SELIS scores by schools and districts within Massachusetts. The conceptual framework for the SELIS was derived from CASEL’s theoretical framework, and premised on a previously validated instrument, the Social and Emotional Competency Assessment (Davidson et al., 2018; Crowder, 2019). Items were included that measured the five core competencies or dimensions: self-awareness (SA), self-management (SM), social awareness (SOC), relationship skills (RSK), and responsible decision-making (RDM). Evidence was provided that supported each aspect of construct validity (content, structural, substantive, generalizability, external, and consequential) for the social and emotional measure. All of the 61-item SELIS survey fit the Rasch model well; the survey exhibited high technical quality. The SELIS scale met the unidimensional assumption of the Rasch model. Students’ dimension scores (SA, SM, SOC, RSK, and RDM) were moderately to strongly correlated with each other indicating that they were distinct conceptually but structurally related by the overarching social and emotional learning construct. The rating scale structure was used by students as intended by the developers and item difficulty hierarchies for each of the dimensions met developers’ *a priori* expectations.

The SELIS survey was reliable at the student and school level. Differential item function analyses indicated that students from different student groups with the same score largely had, within measurement error, the same affirmation level and likely interpretation of most items. These data support the generalizability of the social and emotional skills construct. SELIS scores were, as expected, appropriately related to external measures; for example, the scores were positively related to students’ student- and school- level achievement. The SELIS and dimension measures were responsive at both the student and school level. Students and schools were meaningfully differentiated by their SELIS scores.

Limitations of Study. Unfortunately, due to the pandemic over the last year, the external validity evidence provided is limited in scope. Given the different modes of learning (in-person, all remote, and hybrid) students were participating in throughout the year, using this data to assess the external validity of the SELIS scores is problematic. Going forward, the concurrent relationships with students’ SELIS scores with other behavioral data such as attendance, suspension, and school climate data is needed.

In conclusion, using the data available, the psychometric properties of the SELIS instrument met the basic assumptions of the Rasch-model, namely the items are well-fitting, invariant, and form a unidimensional scale, and showed preliminary evidence to support its external validity.

References

Boone, W. J., Staver, J. R., and Yale, M. S. (2014). *Rasch analysis in the human sciences*, New York: Springer.

Crowder, M. K., Gordon, R. A., Brown, R. D., Davidson, L. A., & Domitrovich, C. E. (2019). Linking social and emotional learning standards to the WCSD Social-Emotional Competency Assessment: A Rasch approach. *School Psychology, 34* (3), 282 – 295.

Davidson, L. A., Crowder, M. K., Gordon, R. A., Domitrovich, C. E., Brown, R. D., & Hayes, B. I. (2018). A continuous improvement approach to social and emotional competency measurement*. Journal of Applied Developmental Psychology. 55*, 93 – 106.

Linacre, J. M. (2021). *A user’s guide to Winsteps, Ministep Rasch-model computer programs: program manual* 5.0.0, Chicago, US: MESA Press.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012 – 1027.

Messick, S. (1995a). Validity of psychological assessment: Validation of inferences from persons’ responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50* (9), 741 – 749.

Messick, S. (1995b). Standards of validity and the validity of standards in performance assessment. *Educational Measurement, Issues and Practice.* 14, 5 – 8.

Smith, A. B. Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology, 8*,

33 – 44.

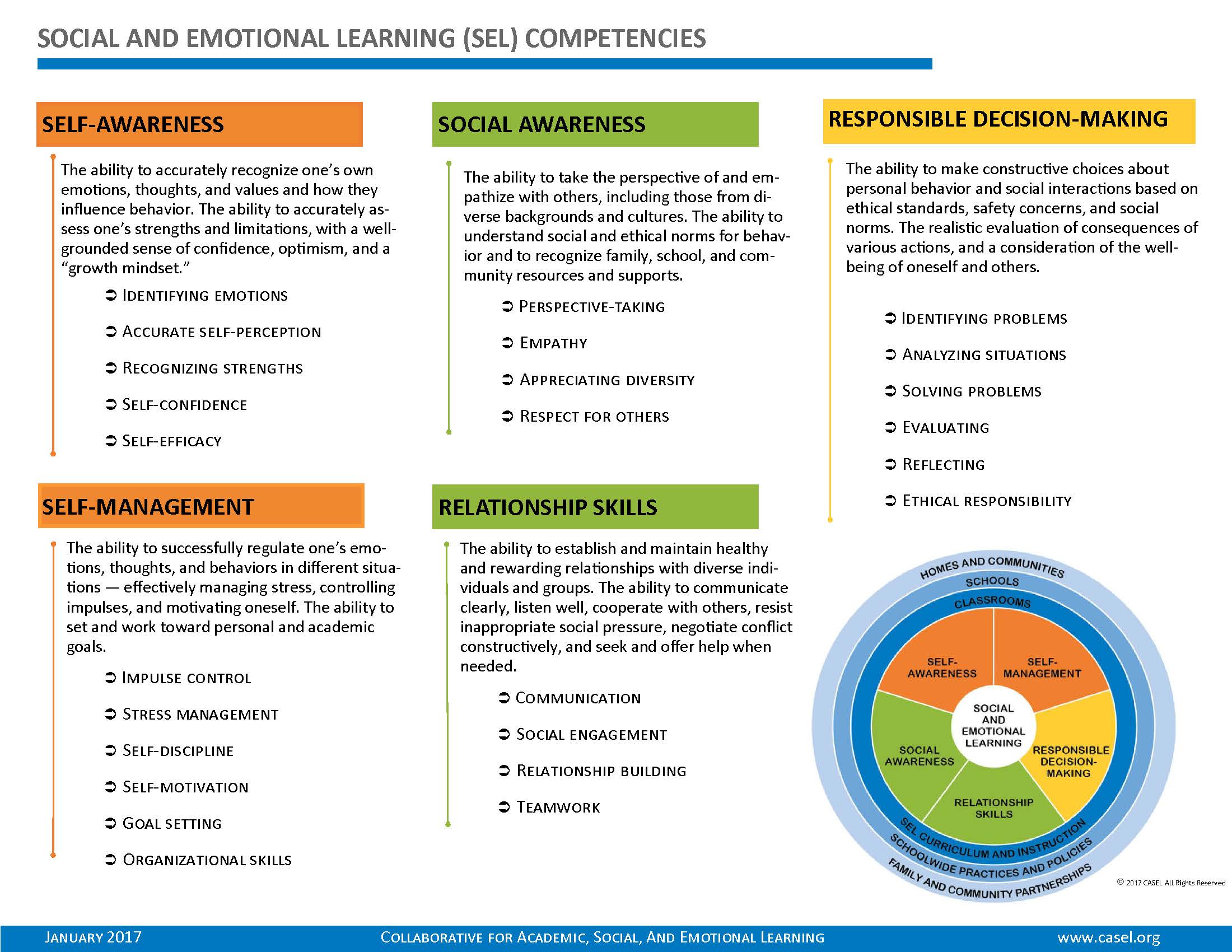
Wolfe, E. W., & Smith, E. V. Jr. (2007a). Instrument development tools and activities for measure validation using Rasch models: Part I – Instrument development tools. *Journal of Applied Measurement*, *8* (1), 97 – 123.

Wolfe, E. W. & Smith Jr., E. V. (2007b). Instrument development tools and activities for measure validation using Rasch models: Part II – Validation activities. *Journal of Applied Measurement*, *8* (2), 204 – 234.

Wright B.D., & Linacre J.M. (1994) Reasonable mean-square fit values. *Rasch Measurement Transactions, 8:3* p.370

Appendix A: Collaborative for Academic and Social and Emotional Learning’s (CASEL) Conceptual Framework

The SELIS survey was based on the **old CASEL framework** which is provided below. Over the summer of 2021, SELIS will be updated to reflect the new CASEL framework that can be found[**here**](https://casel.org/sel-framework/)**.**



Downloaded from [www.casel.org](http://www.casel.org) (2019)

Appendix B: Survey Item Prompts

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No.1** | **Item code** | **Item Prompt: How hard to easy is the following for you?** | **ES2** | **MS2** | **HS2** |
| 11 | SAS1 | Knowing what my strengths are. | X | X | X |
| 21 | SAE2 | Knowing ways to make myself feel better when I'm sad. | X | X | X |
| 31 | SAE3 | Noticing what my body does when I am nervous. | X | X | X |
| 41 | SAE4 | Knowing the emotions I feel. | X | X | X |
| 51 | SAS5 | Knowing when I am wrong in the way I behave. | X | X | X |
| 61 | SAS6 | Knowing how to get better at things that are hard for me to do at school. | X | X | X |
| 71 | SAE7 | Knowing when my mood affects how I treat others. | X | X | X |
| 81 | SAE8 | Knowing ways I calm myself down. | X | X | X |
| 91 | SAS9 | Knowing when I can't control something. | X | X | X |
| 101 | SAE10 | Knowing when my feelings are making it hard for me to focus. | X | X | X |
| 11 | SAE11 | Understanding the difference between how others expect me to feel and how I really feel. |  |  | X |
| 12 | SAS12 | Knowing what I am interested in at school. | X |  |  |
| 13 | SAS13 | Offering an answer even though I am not sure I am right. |  | X |  |
| 14 | SAS14 | Knowing my opinions matter even when others don't listen to them. |  |  | X |
|  |  | Total Number of SA Items | 11 | 11 | 12 |
| 151 | SME1 | Staying calm when I feel stressed. | X | X | X |
| 161 | SME2 | Being patient even when I am really excited. | X | X | X |
| 171 | SMS3 | Working on assignments even when they are hard. | X | X | X |
| 181 | SMS4 | Staying focused in class even when there are distractions. | X | X | X |
| 191 | SMS5 | Planning ahead so I can turn a project in on time. | X | X | X |
| 201 | SMS6 | Being prepared for tests. | X | X | X |
| 211 | SMG7 | Thinking through the steps it will take to reach my goals. | X | X | X |
| 221 | SMG8 | Setting goals for myself. | X | X | X |
| 231 | SMS9 | Finishing my schoolwork without reminders. | X | X | X |
| 241 | SMS10 | Doing my schoolwork even when I do not feel like it. | X | X | X |
| 251 | SMG11 | Finishing tasks even if they are hard for me. | X | X | X |
| 261 | SME12 | Getting through something even when I feel frustrated. | X | X | X |
| 271 | SME13 | Working on things even when I don't like them. | X | X | X |
| 281 | SMG14 | Reaching goals that I set for myself. | X | X | X |
| 29 | SME15 | Trusting at least one adult in school who I can go to if I have a problem. | X | X | X |
| 30 | SME16 | Being careful about what information I share about myself on social media (e.g., TikTok, Facebook, Instagram, etc.). |  | X |  |
| 31 | SMG17 | Catching up on my work when I get behind. | X |  |  |
| 32 | SMG18 | Planning my work when I have multiple assignments due at the same time. |  |  | X |
| Total Number of Self-management (SM) Items | | | 16 | 16 | 16 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Appendix B: Survey Item Prompts continued | | | | | |
| No.1 | Item code | Item Prompt: **How hard to easy is the following for you?** | ES2 | MS2 | HS2 |
| 331 | SOC1 | Knowing what people are feeling by the look on their face. | X | X | X |
| 341 | SOC2 | Knowing how to get help when I'm having trouble with a classmate. | X | X | X |
| 351 | SOC3 | Learning from people with different opinions than me. | X | X | X |
| 361 | SOC4 | Knowing how my actions impact my classmates. | X | X | X |
| 371 | SOC5 | Knowing when someone needs help. | X | X | X |
| 38 | SOC6 | Waiting for other students to finish talking before I speak. | X |  |  |
| 39 | SOC7 | Knowing the difference between bullying someone and joking with someone. |  | X |  |
| 40 | SOC8 | Knowing when to tell my friends how I really feel |  | X |  |
| 41 | SOC9 | Knowing when my tone of voice may cause someone to misunderstand what I intended to say. |  |  | X |
| 42 | SOC10 | Recognizing when I am making someone uncomfortable. |  |  | X |
| Total Number of Social Awareness (SOC) Items | | | 6 | 7 | 7 |
| 431 | RSK1 | Being welcoming to someone I don't usually eat lunch with. | X | X | X |
| 441 | RSK2 | Talking to an adult when I have problems at school. | X | X | X |
| 451 | RSK3 | Getting along well with my classmates. | X | X | X |
| 461 | RSK4 | Sharing what I am feeling with others. | X | X | X |
| 471 | RSK5 | Respecting a classmate's opinions during a disagreement. | X | X | X |
| 481 | RSK6 | Getting along well with my teachers. | X | X | X |
| 49 | RSK7 | Leading a discussion in class. |  | X |  |
| 50 | RSK8 | Asking a classmate to do their fair share of a group project. |  |  | X |
| 51 | RSK9 | Standing up for someone even if they are outside of my friend group. |  |  | X |
| Total Number of Relationship Skills (RSK) Items | | | 6 | 7 | 8 |
| 521 | RDM1 | Helping to make my school a better place. | X | X | X |
| 531 | RDM2 | Knowing when people's actions are "right" or "wrong" (e.g., helpful or harmful) | X | X | X |
| 541 | RDM3 | Thinking of different ways to solve a problem (e.g., a disagreement with another student). | X | X | X |
| 551 | RDM4 | Saying "no" to a friend who wants to break the rules. | X | X | X |
| 561 | RDM5 | Thinking about what might happen before making a decision. | X | X | X |
| 57 | RDM6 | Taking responsibility for my choices. | X |  |  |
| 58 | RDM7 | Knowing when to accept help when it is offered. |  | X |  |
| 59 | RDM8 | Understanding the importance of what I am learning to my future success. |  | X |  |
| 60 | RDM9 | Deciding what courses to take to get me into the college or career I want. |  |  | X |
| 61 | RDM10 | Knowing something is wrong to do even when it happens all the time. |  |  | X |
| Total Number of Responsible Decision-making (RDM) Items | | | 6 | 7 | 7 |
| Total Number of SELIS Survey Items | | | 45 | 48 | 50 |
| 1Items taken or adapted from the Social and Emotional Competency Assessment (SECA), (Crowder et al, 2019; Davidson et al, 2018) | | | | | |
| 2ES: Elementary (G3 - G5) items; MS: Middle school (G6 - G8); HS: High school (G9-G12) | | | | | |

**Appendix C: Survey Specifications**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Competency** | **Domain** | **G3 – G5**  **Items1** | **G6 – G8**  **Items1** | **G9 – G11**  **Items1** | **Total2** |
| **Self-awareness (SA)** | SA emotion (SAE) | 6+0 | 6+0 | 6+1 | 7 |
| SA self-concept (SAS) | 4+1 | 4+1 | 4+1 | 7 |
| SA Subtotal |  | 11 | 11 | 12 | 14 |
| **Self-Management (SM)** | SM emotion (SME) | 5+0 | 5+1 | 5+0 | 6 |
| SM schoolwork (SMS) | 6+0 | 6+0 | 6+0 | 6 |
| SM goals (SMG) | 4+1 | 4+0 | 4+1 | 6 |
| SM Subtotal |  | 16 | 16 | 16 | 18 |
| **Social-awareness (SOC)** | Social awareness | 5+1 | 5+2 | 5+2 | 10 |
| **Relationship skills (RSK)** | Relationship skills | 6+0 | 6+1 | 6+2 | 9 |
| **Responsible Decision-making (RDM)** | Responsible decision-making | 5+1 | 5+2 | 5+2 | 10 |
| Subtotal |  | 18 | 21 | 22 | 29 |
| **TOTAL**  1The number before the plus sign is the number of common items that appear across the three grade-level survey forms; the number after the plus sign is the number of items unique to the grade-level survey form; 2Common items are only counted once to derive the total. |  | 45 | 48 | 50 | 61 |

Appendix D: Intra-class correlations for SELIS data

|  |  |  |
| --- | --- | --- |
| SE Construct/Dimension | School (N = 82) | Teacher (N = 515)1 |
| Social and Emotional | .015 | .041 |
| Self-awareness | .012 | .030 |
| Self-management | .013 | .038 |
| Social awareness | .007 | .024 |
| Relationship Skills | .021 | .033 |
| Responsible Decision-making | .016 | .037 |

1Two districts did not supply teacher-level information in their rosters

Appendix E: Linear Regression Results

E1: Predicting SE construct with English Language Arts & Literacy scaled scores (escaleds) using student characteristics as covariates

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model Summary | | | | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error | Change Statistics | | | | | Durbin-Watson |
| R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .095a | .009 | .009 | .8827 | .009 | 96.010 | 1 | 10448 | .000 |  |
| 2 | .141b | .020 | .019 | .8781 | .011 | 22.866 | 5 | 10443 | .000 | 1.921 |
| a. Predictors: (Constant), Zscore(escaleds) | | | | | | | | | | |
| b. Predictors: (Constant), Zscore(escaleds), racebi, gender, EL, ecodis, SWD | | | | | | | | | | |
| c. Dependent Variable: ZDSEms | | | | | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model  Adjusted R square: 0.019 | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | -.006 | .009 |  | -.749 | .454 |
| Zscore(escaleds) | .085 | .009 | .095 | 9.798 | .000 |
| 2 | (Constant) | .034 | .016 |  | 2.200 | .028 |
| Zscore(escaleds) | .054 | .010 | .061 | 5.343 | .000 |
| gender | .090 | .017 | .051 | 5.162 | .000 |
| racebi | -.065 | .019 | -.035 | -3.443 | .001 |
| ecodis | -.117 | .019 | -.064 | -6.079 | .000 |
| EL | -.012 | .048 | -.003 | -.253 | .800 |
| SWD | -.117 | .025 | -.049 | -4.640 | .000 |

Dependent variable: ZDSEms; male, students of color (racebi), economically disadvantaged (ecodis), English learner (EL), and students with disabilities (SWD) coded 1.

E2: Predicting SE construct with English Language Arts & Literacy student growth percentile scores (esgp) using student characteristics as covariates

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model Summary | | | | | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error | Change Statistics | | | | | Durbin-Watson |
| R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .033a | .001 | .001 | .8698 | .001 | 7.668 | 1 | 7130 | .006 |  |
| 2 | .124b | .015 | .015 | .8638 | .014 | 20.798 | 5 | 7125 | .000 | 1.907 |
| a. Predictors: (Constant), Zscore(esgp) | | | | | | | | | | | |
| b. Predictors: (Constant), Zscore(esgp), racebi, SWD, gender, EL, ecodis | | | | | | | | | | | |
| c. Dependent Variable: ZDSEms | | | | | | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model  Adjusted R square: 0.014 | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | .003 | .010 |  | .276 | .783 |
| Zscore(esgp) | .029 | .010 | .033 | 2.769 | .006 |
| 2 | (Constant) | .036 | .017 |  | 2.072 | .038 |
| Zscore(esgp) | .028 | .010 | .032 | 2.700 | .007 |
| gender | .106 | .021 | .061 | 5.125 | .000 |
| racebi | -.077 | .023 | -.042 | -3.402 | .001 |
| ecodis | -.135 | .023 | -.073 | -5.847 | .000 |
| EL | .024 | .068 | .004 | .353 | .724 |
| SWD | -.096 | .029 | -.040 | -3.308 | .001 |

E3: Predicting SE construct with Mathematics scaled scores (mscaleds) using student characteristics as covariates

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model Summary | | | | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error | Change Statistics | | | | | Durbin-Watson |
| R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .106a | .011 | .011 | .8811 | .011 | 119.480 | 1 | 10496 | .000 |  |
| 2 | .141b | .020 | .019 | .8774 | .009 | 18.435 | 5 | 10491 | .000 | 1.923 |
| a. Predictors: (Constant), Zscore(mscaleds) | | | | | | | | | | |
| b. Predictors: (Constant), Zscore(mscaleds), gender, racebi, EL, ecodis, SWD | | | | | | | | | | |
| c. Dependent Variable: ZDSEms | | | | | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model  Adjusted R square: .019 | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | -.007 | .009 |  | -.816 | .415 |
| Zscore(mscaleds) | .094 | .009 | .106 | 10.931 | .000 |
| 2 | (Constant) | .042 | .015 |  | 2.789 | .005 |
| Zscore(mscaleds) | .055 | .010 | .062 | 5.425 | .000 |
| gender | .070 | .017 | .039 | 4.057 | .000 |
| racebi | -.066 | .019 | -.036 | -3.485 | .001 |
| ecodis | -.114 | .019 | -.062 | -5.871 | .000 |
| EL | -.007 | .046 | -.002 | -.156 | .876 |
| SWD | -.115 | .026 | -.048 | -4.513 | .000 |

E4: Predicting SE construct with Mathematics student growth percentile scores (msgp) using student characteristics as covariates

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summary** | | | | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error | Change Statistics | | | | | Durbin-Watson |
| R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .039a | .002 | .001 | .8694 | .002 | 11.093 | 1 | 7131 | .001 |  |
| 2 | .125b | .016 | .015 | .8636 | .014 | 20.237 | 5 | 7126 | .000 | 1.913 |
| a. Predictors: (Constant), Zscore(msgp) | | | | | | | | | | |
| b. Predictors: (Constant), Zscore(msgp), racebi, gender, SWD, EL, ecodis | | | | | | | | | | |
| c. Dependent Variable: ZDSEms | | | | | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model  Adjusted R square: 0.014 | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | .003 | .010 |  | .281 | .811 |
| Zscore(msgp) | .034 | .010 | .039 | 3.331 | .001 |
| 2 | (Constant) | .036 | .017 |  | 2.110 | .031 |
| Zscore(msgp) | .028 | .010 | .032 | 2.730 | .008 |
| gender | .103 | .021 | .059 | 5.011 | .000 |
| racebi | -.075 | .023 | -.041 | -3.306 | .001 |
| ecodis | -.133 | .023 | -.071 | -5.744 | .000 |
| EL | .015 | .067 | .003 | .219 | .826 |
| SWD | -.101 | .029 | -.042 | -3.457 | .001 |

1. Many items were taken as is or adapted from the Social-Emotional Competency Assessment (Chowder M. K. et al. (2019). Linking social and emotional learning standards to the WCSD Social-Emotional Competency Assessment: A Rasch approach. *School Psychology, 34*(3*),* 281–295)*.* [↑](#footnote-ref-2)