

Multilevel Reliability for Social and Emotional Learning Measures

Carlos Chavez, Michael C. Rodriguez, Kyle Nickodem, Kory Vue, Ta Do
University of Minnesota

Minnesota Youth Development Research Group
www.mnydr.org

September 2020

Paper presented at the annual meeting of the
National Council on Measurement in Education

Citation:

Chávez, C., Rodriguez, M.C., Nickodem, K., Vue, K., & Do, T. (2020, September).
Multilevel reliability of social and emotional learning measures [Paper
presentation]. National Council on Measurement in Education Annual Meeting.
<https://conservancy.umn.edu/handle/11299/194886>

Multilevel Reliability for Social and Emotional Learning Measures Introduction

Background

Despite a growing nationwide interest in social and emotional learning (SEL), including the usage of SEL measures for accountability purposes as allowed by the Every Student Succeeds Act [ESSA] (ESSA, 2016), there is little technical evidence for the use of SEL measures at the school level. As education data are often hierarchically structured (i.e. students nested within schools), it is important to consider multiple levels to support the interpretation and use of SEL measures (AERA, APA, & NCME, 2014). One method to support the interpretation of SEL measures is to estimate score reliability at the person and the school level. Reliability is often estimated as the ratio of true score variance over total score variance. One commonly reported metric of reliability is coefficient alpha; however, it may be inappropriate to report alpha as its stringent assumptions are rarely met (e.g., essential tau equivalency). Alternate estimates exist, namely omega and maximal, or H, reliability, that are less strict and allow for congeneric measurement models.

Perspectives

Social and emotional learning (SEL) is core component underlying our research. The ecologies of youth development involve complex networks of social structures including family, peers, school, and community, that impact personal, academic, and social outcomes (Bronfenbrenner, 1994; Lerner, 2004). When students are equipped with the supports from multiple relationships and contexts, their development is positively influenced. Researchers and media tend to focus on the negative effects of discrimination, associated with at-risk conditions or personal and community-based deficits, without providing a means of support to interrupt negative outcomes. To focus on positive outcomes, we employ a positive youth development (PYD) orientation, an asset-oriented perspective (Damon, 2004). Our research is inspired by several key PYD principles: (a) youth have an inherent capacity for and are key actors in their own positive development; (b) positive development is enabled through multiple relationships; (c) all youth benefit from positive opportunities; and (d) community is a critical delivery system (Benson et al., 2006).

One interest of this paper is to add to the potential evidence of noncognitive measures by accounting for group differences in the factor structure of the construct. As education data are often hierarchically structured, the factor structure should also be considered to identify any potential dependencies within the data. To date, there is little work that accounts for group differences when providing internal structure and construct related evidence. Dyer et al. (2005) made note of the importance in accounting for nested data structure when providing evidence on the factor structure of measures. However, more recent work by Ahn et al. (2019) and Loeb et al. (2019) provide mixed evidence on the utility of a multilevel factor structure for social emotional learning measures.

The objective of this study is to add to the interpretation of SEL measures by accounting for the variation between and within schools. To guide our objective, we propose the following research question:

To what extent does score reliability for SEL measures differ at person level compared to school level?

Methods

Data Source

The data come from a 2016 administration of a triennial statewide survey (blinded for review, 2016, 2017) that was administered to 168,733 youth in grades 5, 8, 9, and 11. This survey is a collaborate effort from the Minnesota Departments of Education, Health, Human Services, and Public Safety. The survey is intended to monitor trends in students' habits, experiences, and beliefs.

Measures

Using the developmental asset framework, we derived six SEL measures including: Commitment to Learning (CtL), Positive Identity and Outlook (PIO), Social Competence (SC), Empowerment (EM), Family Community Support (FCS), and Teacher School Support (TSS). Developmental skills are represented by CtL, PI, and SC. Developmental supports are represented by EM, FCS, and TSS. For this paper we used the measures representing developmental skills. CtL consists of six items and is defined as 'caring about doing well in school, paying attention in class, going to class prepared, interested in learning, finding school learning useful, and being a student is an important part of who I am.' PI consists of six items and is defined as 'having a sense of control of one's life, feeling good about self and future, dealing well with disappointment and life's challenges, and thinking about one's purpose in life.' SC consists of eight items and is defined as 'saying no to dangerous/unhealthy things, building friendships, expressing feelings appropriately, planning ahead and making good choices, resisting bad influences, resolving conflicts without violence, accepting differences in others, and recognizing the needs and feelings of others.

Initial evidence for the use of these measures come from confirmatory factor analysis to support construct related evidence and content related evidence based on the work of Benson et al. (2006) and the Search Institute (2013). Differential item functioning analyses by race & ethnicity, assigned sex at birth, and grade was also done in Winstep v 3.92 (Linacre, 2016). Common guidelines for assessing model fit of these measures were used, which include adequate fit indices of Root Mean Square Error Approximation (RMSEA) values less than .05, comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) values are greater than .95 (Brown, 2015; Kline, 2011), and standardized factor loadings are .40 or higher (Brown, 2015).

We used three measures of developmental skills (CtL, PIO, and SC) and estimated both single-level and multi-level alpha, omega, and H reliabilities in Mplus Version 8.0 (Múthen & Múthen, 2018). This paper applies the methods developed by Geldhof et al. (2014).

Confirmatory factory analysis (CFA) was conducted to estimate alpha, omega, and H reliabilities at the single level, and a multilevel CFA (MCFA) was used to estimate multilevel reliabilities.

Raw scores were clustered by school ID to explain variation within and between schools. Although the items used in our measures are categorical, due to software limitations the variables were treated as continuous.

Results

Overall, reliability estimates improved when accounting for variation between schools. Additionally, maximal reliability, reported the highest estimates out of the three. However, it should be noted that H is often an inconsistent estimator of reliability. Omega reliability generally had higher estimates of reliability than alpha. This finding was consistent with both single-level and multilevel estimations. The intraclass correlations (ICC) for our measures, see Table 1, were all below .10, which in previous research, low ICCs present a positive bias in the estimation of reliability given a high number of clusters at the between level (Geldhof, 2014). This may account for the large differences between reliability estimates, and in particular H. Tables 2 and 3 provide a look at the single-level and multilevel reliability estimates. CtL and SC showed only marginal differences between reliability estimates, whereas PIO had estimates that ranged between .762 and .872. For the between-cluster estimates, the lowest value was an alpha value of .712 for CtL whereas PIO had an estimated H value of .996.

The student-level estimates were similar to Rasch model reliabilities reported in technical reports for these measures. This study explored the consistency between student-level and school-level score reliabilities. Since the measures are intended for use at the group (school) level, the multilevel estimates are more appropriate.

Table 1
Intraclass Correlations for Multilevel Alpha, Omega, and H of the Developmental Skills Measures

Measure	Variable	ICC
Commitment to Learning	Y18	.033
	Y19	.034
	Y20r	.024
	Y21a	.023
	Y21b	.113
	Y21c	.058
Positive Identity and Outlook	Y60a	.034
	Y60b	.051
	Y60f	.041
	Y60g	.018
	Y60h	.024
	Y60n	.012
Social Competence	Y60c	.032
	Y60d	.028
	Y60e	.034
	Y60i	.031
	Y60j	.050
	Y60k	.040
	Y60m	.037
	Y60q	.027

Table 2
Single-Level Reliability Estimates for SEL Measures

Measures	Reliability	Estimate	SE
Commitment to Learning	Alpha	.677	.001
	Omega	.699	.001
	H	.750	.001
Positive Identity & Outlook	Alpha	.762	.001
	Omega	.841	.001
	H	.872	.001
Social Competence	Alpha	.839	.001
	Omega	.841	.001
	H	.850	.001

Note. These reliability estimates are similar to the person level Rasch score reliabilities, CtL (.70), PIO (.79), and SC (.79).

Table 3
Multi-Level Reliability Estimates for SEL Measures

Measure	Reliability	Level	Estimate	SE
Commitment to Learning	Alpha	Within	.676	.002
		Between	.712	.015
	Omega	Within	.696	.002
		Between	.889	.008
	H	Within	.743	.002
		Between	.968	.006
Positive Identity and Outlook	Alpha	Within	.830	.001
		Between	.930	.005
	Omega	Within	.837	.001
		Between	.960	.001
	H	Within	.868	.003
		Between	.996	.002
Social Competence	Alpha	Within	.832	.001
		Between	.973	.002
	Omega	Within	.839	.001
		Between	.841	.005
	H	Within	.844	.001
		Between	.988	.002

Significance

As the demand for SEL measures increases across the nation, it is important to contribute all possible evidence for the interpretation and use of these measures. Despite the increased demand, there has been little psychometric evidence that can be used for interpreting SEL measures at the school level. The results of our study have shown that raw score measures of SEL are more reliable at the between-cluster level than at the single-level. However, we did run into the limitation of having to treat categorical variables as continuous. Nevertheless, the results of this study may provide new evidence for the interpretation and use of SEL measures for schools.

References

- American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (2014). *Standards for educational and psychological testing*. AERA.
- Ahn, I., Patrick, H., Chiu, M.M., & Levesque-Bristol, C. (2019). Measuring teacher practices that support student motivation: Examining the factor structure of the Teacher as Social Context Questionnaire using multilevel factor analyses. *Journal of Psychoeducational Assessment*, 37(6), 743–756. <https://doi.org/10.1177/0734282918791655>
- Benson, P.L., Scales, P.C., Hamilton, S.F., & Sesma, A. (2007). Positive youth development: Theory, research, and applications. In W. Damon & R.M. Lerner (Eds.), *Handbook of child psychology* (6th ed., pp. 894-941). John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470147658.chpsy0116>
- Brown, T. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.
- Dyer, N.G., Hanges, P.J., & Hall, R.J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly*, 16(1), 149–167. <https://doi.org/10.1016/j.leaqua.2004.09.009>
- ESSA (2015). Every Student Succeeds Act of 2015, Pub. L. No. 114-95 § 114 Stat. 1177.
- Geldhof, G.J., Preacher, K.J., & Zyphur, M.J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, 19(1), 72–91. <https://doi.org/10.1037/a0032138>
- Linacre, J.M. (2016). Winsteps® (Version 3.92.0) [Computer Software]. Winsteps.com. <http://www.winsteps.com/>
- Loeb, S., Christian, M. S., Hough, H., Meyer, R.H., Rice, A.B., & West, M.R. (2019). School differences in social–emotional learning gains: Findings from the first large-scale panel survey of students. *Journal of Educational and Behavioral Statistics*, 44(5), 507– 542. <https://doi.org/10.3102/1076998619845162>
- MN Department of Education (2017). *Minnesota Student Survey*. <https://education.mn.gov/MDE/dse/health/mss/>
- Muthén, L.K., & Muthén, B.O. (2012). Mplus. (Version 7). [Software program].
- MN Department of Education. (2018a). *SEL implementation guidance*. <https://education.mn.gov/MDE/dse/safe/clim/social/imp/>
- Rodriguez, M.C. (2017). *Technical report on developmental skills, supports, & challenges from the 2013-2016 Minnesota Student Survey*. <https://conservancy.umn.edu/handle/11299/195197>
- Search Institute. (2013). Developmental Assets Profile: Technical summary. <http://www.search-institute.org/surveys/dap>