Assessment in Education: Principles, Policy & Practice

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/caie20

Building useful, web-based educational assessment tools for students, with students: a demonstration with the school climate walkthrough

Jessica D. Hoffmann, Rachel Baumsteiger, Jennifer Seibyl, Elinor Hills, Christina Bradley, Christina Cipriano & Marc A. Brackett

To cite this article: Jessica D. Hoffmann, Rachel Baumsteiger, Jennifer Seibyl, Elinor Hills, Christina Bradley, Christina Cipriano & Marc A. Brackett (2022): Building useful, web-based educational assessment tools for students, with students: a demonstration with the school climate walkthrough, Assessment in Education: Principles, Policy & Practice, DOI: 10.1080/0969594X.2022.2056576

To link to this article: https://doi.org/10.1080/0969594X.2022.2056576

View supplementary material

Published online: 28 Mar 2022.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=caie20

ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE
https://doi.org/10.1080/0969594X.2022.2056576

Routledge
Taylor & Francis Group

Check for updates

Building useful, web-based educational assessment tools for students, with students: a demonstration with the school climate walkthrough

Jessica D. Hoffmann, Rachel Baumsteiger, Jennifer Seibyl, Elinor Hills, Christina Bradley, Christina Cipriano and Marc A. Brackett

Yale Center for Emotional Intelligence, Yale University, New Haven, CT, USA

ABSTRACT

Practical educational assessment tools for adolescents champion student voice, use technology to enhance engagement, highlight discrepancies in school experience, and provide actionable feedback. We report a series of studies that detail an iterative process for developing a new school climate assessment tool: (1) item generation that centres student voice, (2) the design of a web-based app, (3) item revisions informed by student and educator feedback, and (4) the identification and confirmation of the underlying factor structure of the assessment tool. The resulting School Climate Walkthrough provides scores on nine dimensions of school climate and 73 additional observational items. The web-based application produces instantaneous reports that display systemic differences in how various student demographic groups experience school. This process can guide future research in building the next generation of educational assessments for adolescents and disrupting harmful or exclusionary school practices.

ARTICLE HISTORY
Received 16 December 2020
Accepted 14 March 2022

KEYWORDS
School climate; adolescents; measure; scale development; education; youth empowerment

A goal for the field of educational assessment is to maximise the utility of assessment tools (Bennett, 2018; Pellegrino, 2020). One standard metric is a measure's psychometric properties such as internal consistency, test-retest reliability, and convergent validity (Boateng et al., 2018; Rust & Golombok, 2014). Yet, while these measurement considerations are essential, they are insufficient. Qualities such as strategic use of technology, a focus on individuals, and the provision of actionable feedback have been part of the discussion for decades, yet most assessments of school climate do not meet these standards (Olsen et al., 2018). To begin, we articulate why these features are integral, then describe a process for developing an assessment tool that achieves these aims, illustrated by the development of a school climate assessment, The School Climate Walkthrough Tool.

CONTACT Jessica D. Hoffmann Jessica.hoffmann@yale.edu 350 George Street, New Haven 06520, CT, USA
Twitter: @JADHoffmann, @rachbaumsteiger, @drchriscip, @marcbrackett
Supplemental data for this article can be accessed here
© 2022 Informa UK Limited, trading as Taylor & Francis Group

Strategic use of technology

Self-report surveys are a quick, inexpensive method for gathering student responses; however, low response rates, survey fatigue, and response biases can detract from accuracy and representativeness (Cornell et al., 2012). These problems are especially common for surveys more than 20 minutes (Revilla & Ochoa, 2017) and can lead to missing data, or cause participants to rush, providing less thoughtful responses or selecting answers at random. Studies indicate that random responding is common in adolescents, with rates estimated between 26% and 60% among college students (Wise, 2006; Berry et al., 1992, respectively), and approximately 40% among middle school students (Osborne & Blanchard, 2011). Even small numbers of invalid responses can significantly alter study results (e.g. Cornell & Loper, 1998; M. Furlong et al., 2004). Yet, strategies aimed at reducing survey length, such as only retaining items that maximise internal consistency, raise new challenges (e.g. factor structure issues; Stanton et al., 2002). One solution is to enhance student engagement by strategically using the many features made available by modern technology.

Contemporary adolescent has already been socialised to access and interact with web-based applications as a means of participation in their learning and relationships (Family Foundation, 2010; Rideout & Robb, 2019), and tend to favour completing surveys on electronic devices over paper-and-pencil formats (Dubad et al., 2018). Plus, electronic platforms afford many advantages that can make surveys feel more personalised, efficient, and fun. For example, there are more options for survey flow (piping text, branching rules), and features to optimise accessibility (screen readers, text presentation). Electronic platforms allow for the collection of metadata (e.g. start time, time spent on a page), opportunities to animate and gamify the survey – which can make surveys feel more enjoyable (e.g. Mavletova, 2014); and, automatic aggregation of responses, which facilitates instantaneous scoring (DeRosier & Thomas, 2018).

A focus on the individual

Historically, most educational assessments use variable-centred approaches, which focus on group averages, such as student's perception of safety at school or incidence rates of bullying. Yet, averages do not acknowledge the range of responses within a sample which dilute less common responses (Clawar et al., 2018) and conceal discrepancies among subgroups (M. J. Furlong et al., 2005). For example, Varjas et al. (2009) found that male students report more physical and verbal victimisation at school while female students report more relational victimisation. A promising trend in educational measurement is to examine interpersonal variance such as inconsistencies in disciplinary practices for students of differing races (Bottiani et al., 2017). Statistical techniques such as latent class analysis allow for more within-person analysis, such as investigating the experiences of students in specific demographic groups (Romero & O'Malley, 2020), and the study of experiences of school climate among students who are chronically absent (Van Eck et al., 2017). Following this approach, assessments can display results by demographic group and emphasise the range of responses rather than a single average. Of course, this means assessments must be engaging and accessible so as to acquire large, representative samples necessary for subgroup comparisons while preserving individual anonymity.

Meaningful, actionable feedback

Assessments with systematic validity are those that facilitate improvements in the qualities that they are designed to measure (Frederiksen & Collins, 1989). Thus, a school climate assessment would have systematic validity if it led to improvements in school climate. There are many calls for assessments to provide more support to the people who complete them (U.S. Education Council, 2019).

One clear direction for making assessments more supportive is to provide easily interpretable reports that offer guidance on how to improve in the future (Assessment Work Group, 2019). Assessment developers can collaborate with students, teachers, and school leaders to determine what makes a useful report and what type of suggestions will be realistic and relevant. Helpful feedback is also timely. The student population of a school shifts each year as a result of graduations; therefore, assessment developers should provide reports quickly to ensure that the results are relevant, such as through automated website features or apps with scoring capabilities.

Placing students at the center from the start

The majority of educational assessments are selected and administered by school leaders and reports are provided directly to them, excluding the students in data interpretation and decisions about next steps. Naturally, this can lead students to provide survey responses with little thought (Cronbach, 1950) due to lack of motivation, a rebellious reaction to being measured, or disinterest (Berry et al., 1992; Meier, 1994; Wise, 2006), the consequences of which were discussed previously. Another student response can be disengaging from the survey out of a sense of futility that nothing will be done with the results (Cornell University, 2012). A study of student evaluations of teaching found students were more likely to be honest when they thought the evaluations were an effective measure of the course quality, and when the results would be used to benefit future students (McClain et al., 2018). Assessment tools can ensure they have the questions students find most relevant by including students in the writing of those items, which can also increase confidence that the data will be interpreted and used well.

In sum, there is broad consensus that educational assessments should leverage technology, examine individual differences, and generate meaningful reports. We add that, to ensure the assessment tool is engaging and the reports are meaningful, students must be key decision-makers in the process.

The present research

The goal of the current work was to develop a measure of school climate ('The Walkthrough'). We begin with a brief review of the school climate assessment literature, highlighting how the features enumerated above–engaging technology, an individual focus, and actionable feedback–are relevant to this construct. We then describe the development of the Walkthrough. Student feedback is woven throughout the studies' methods, and we note the key outcomes that this process yielded. Finally, we discuss how this process could be applied to educational assessment development more broadly.

School climate

School climate refers to a school's norms and values, teaching practices, and quality of relationships (Cohen et al., 2009; M.-T. Wang & Degol, 2016). At the most basic level, students who feel safe and supported at school are more likely to attend, to focus, and therefore to learn. School climate is associated with academic outcomes such as student engagement (Fatou & Kubiszewski, 2018) and achievement (Bear et al., 2011; Thapa et al., 2013; Zullig et al., 2010) and with social behaviours at school such as decreased bullying (Bandyopadhyay et al., 2009; Loukas et al., 2006; M. Wang & Dishion, 2012).

Recently, Marracini and colleagues (2020) called for advancing the existing work on school climate to enhance measurement utility for informing intervention selection, and supporting educational equity by disaggregating demographic characteristics, incorporating diverse samples, and facilitating educational goal-setting. Their call closely aligns with the three key features covered in our introduction: engaging technology, an individual focus, and actionable feedback. School climate is not just about the school, but also the interactions among individuals and the school. Students attending the same school can have vastly different experiences (Gage et al., 2016; M.-T. Wang & Degol, 2016; Romero & O'Malley, 2020). In fact, studies using multilevel modelling indicate that the majority of variance in perceptions of school climate reflects differences across individuals, although a significant portion of variance also reflects differences across schools (e.g. Griffith, 1997; Ramsey et al., 2016). Measuring individual variation results in a more accurate assessment of school climate and provides directions for supporting students whose voices might otherwise have been lost. Group differences are especially important in light of social inequities within the U.S. educational system (U.S. Department of Education, 2020).

Recent systematic reviews illustrate that each of the features described above are included in some of the existing and emergent assessment tools (Grazia & Molinari, 2021; Lenz et al., 2021; M.-T. Wang & Degol, 2016; Marraccini et al., 2020; Olsen et al., 2018), but tend to focus on one priority or another. For example, Olsen and colleagues (2018) identified four school climate assessments with strong psychometric properties, but note that none of these are freely accessible. The goal of our research was to develop an assessment that would bring together all of these strengths into a single tool.

Pilot study 1: student focus groups

The purpose of this pilot was to ask students to generate a list of items that represent theoretical dimensions of school climate. Six 30-minute focus groups were conducted with 72 students at three public high schools (12 students per group): two New York City schools (Schools A and B) and a suburban Connecticut school (School C), each selected to provide their unique perspective. Groups at Schools A and C were approximately equal in male and female composition, while groups at School B were 96% female, proportional to the percent of the student population who identify as female at each particular school. Race/ethnicity and grade information was not collected from individuals in the focus groups, though we had asked educators to gather representative groups for the sessions. Demographic information on the distinct makeup of the schools A, B, and C is provided in Table 1.

Table 1. Demographic information for schools in student focus groups and pilot studies.

	School A	School B	School C	School D	School E	School F
Study	Pilot Study 1	Pilot Study 1	Pilot Study 1	Pilot Study 3	Study 2	Study 2
School Size	518	1606	1828	1577	874	1455
School Type	Public	Public	Public	Public	Public	Public
Locale	Urban	Urban	Suburban	Urban	Rural	Suburban
Free/reduced price lunch	25%	83%	1%	61%	66%	10%
% Male	52%	10%	52%	47%	53%	50%
Grades served	9–12	9–12	9–12	9–12	9–12	9–12
Race/Ethnicity						
African American or Black	2%	30%	2%	38%	5%	4%
American Indian or Alaska Native	1%	1%	0%	1%	<1%	<1%
Asian or Asian-American	19%	4%	6%	3%	6%	15%
Biracial or Multiracial	8%	1%	3%	1%	3%	4%
Hispanic or Latin	12%	57%	6%	43%	30%	2%
Native Hawaiian or Pacific Islander	0%	1%	0%	<1%	<1%	<1%
White or Caucasian	58%	5%	83%	14%	56%	75%

Note. School statistics were gathered from the National Center for Educational Statistics (NCES) database.

[insert Table 1 about here]

After providing verbal assent, students were asked to brainstorm observable indicators of positive or negative school climate. We asked students to consider each domain of school climate outlined by the National School Climate Council (National School Climate Council, 2007), including safety, teaching and learning, relationships, and institutional environment. For example, we asked, 'What are some of the ways you can observe if your school respects diversity?' In the spirit of a 'walkthrough' we then asked students to visualise different locations of the school including the entrances, hallways and stairs, classrooms, cafeteria, bathrooms, and school grounds.

Each focus group yielded new ideas. For example, in School A, students expressed stress related to overcrowding and issues of seniority with using elevators, as their school occupied ten floors of a city building. We also learned that students rarely used the cafeteria and preferred to purchase lunch in the city. In School B, several students shared that their principal was famous for his ability to greet all 1,600 by name each morning in the lobby, and personally worked to ensure that every student who wanted a summer internship received one. In School C, we learned that the school had gone years without a physical altercation, a streak recently broken by perceived discrimination by political affiliation. We share these examples here not to suggest that these particular students' opinions of their school would have been agreed to unanimously by the student body, but rather to provide a window into the kinds of observations students tended to use as 'evidence' of a positive or negative climate. Students' suggestions were combined, with slight wording changes to improve interpretability and similar items collapsed to reduce redundancy. This process led to the generation of 100 student-generated items.

Pilot study 2: feedback from expert advisors

Expert advisors reviewed student-generated items with three criteria in mind: whether items were (1) easy to understand – using a sixth grade vocabulary or lower and simple sentence structure, (2) comprehensive – covering all major locations of a school and domains of school climate, and (3) universal – applying as broadly as possible across school types and sizes, and which have similar connotations for all students. For example,

reference to metal detectors would only apply in some schools, as would naming specific sports or specific courses. Expert advisors included 45 employees of the [authors' institution, masked for review] who have relevant experience, including: (1) former middle and high school educators, principals and superintendents; (2) social, school, and clinical psychologists; (3) experts in social and emotional learning; and (4) experts in assessment development. Experts spent approximately one hour providing notes and suggestions. These suggestions included the identification of items that were not clearly positive or negative. For example, 'Adults monitor the bathrooms for bullying behavior' could indicate a bullying problem or that adults are working to keep all students safe. Such items were reworded or removed. Experts also noted places in which items were double-barrelled and should be divided into two, and places in which jargon should be removed or clarified. The revised list included 87 items.

Pilot study 3: pilot testing

A group of 26 students from an urban public high school in Connecticut (School D) were recruited to pilot the items. Students completed the survey anonymously and we did not collect information on their age, grade, gender, or race/ethnicity. Again, educators were asked to enlist a representative group of students, and the demographic information for the school is presented in Table 1.

Of the 87 items representing the NSCC domains of school climate, the first 59 were observable indices of each domain across six school locations (e.g. 'I saw people wearing school colors and gear today', and 'I saw students getting teased or put down in the hallways, stairs, and/or elevators today'). Because we were only interested in whether participants observed each phenomenon or not (rather than the extent to which they observed them), these items were rated on a dichotomous scale with the response options of 0 (*Not observed*) or 1 (*Observed*; see recommendations by Boateng et al., 2018; Krosnick & Presser, 2010). We also provided a 'Not Applicable' option for cases in which the activity (e.g. lunch) or location (e.g. elevators) were not relevant for that particular school. The remaining 26 items were designed to capture student's overall general impressions of each domain of school climate not tied to a specific location (e.g. 'My school has clear rules' and 'My school makes an effort to ensure that all people from all backgrounds are accepted'). These items were rated on a 4-point Likert scale from 1 (*Strongly disagree*) to 4 (*Strongly agree*).

Students were asked to complete the observable items throughout the school day as they moved through different locations, to complete the general section at the end of the day, and to write any feedback or questions on the survey. Students' feedback included confusion around several terms (e.g. 'gender identity' and 'equitable'). These items were reworded. Other feedback indicated that the remaining items were understandable and relevant. The revised items were used in the next study.

Study 1: refining items and identifying factors

Our primary goal was to identify whether the Walkthrough items represented the theorised domains of school climate (National School Climate Council, 2007) and the six school locations that guided item generation in Pilot Study 1. Additionally, we wanted the final assessment tool to have good discriminant validity, meaning that it can

differentiate perceptions of school climate across students and schools. Therefore, we evaluated item variances for the purpose of eliminating items that did not help to distinguish among individuals and schools.

Method

Students received a paper-and-pencil version of the 87 revised items described in Pilot Study 3. Students were instructed to complete the observable items throughout the school day and complete the general section at the end of the day. Surveys were returned by mail to the researchers, who provided schools with a report of the findings within two weeks.

Participants

The sample included 507 students from nine secondary schools in California, New York, Connecticut, New Hampshire, and Virginia. This sample size satisfies the recommendation of having at least 5 participants for each item (5 x 87 items = 435) and a minimum number of 300 participants for an exploratory factor analysis (DeVellis, 2017). As with the pilot studies, data was collected anonymously, and we did not gather demographic information from the students. Table 2 provides the demographic information for the nine schools; schools were selected to achieve a diverse sample of students in suburban, urban, and rural locations, attending private and public school, with a range of socio-economic statuses and varying racial compositions.

[insert Table 2 about here]

Data analyses

We performed an exploratory factor analysis with fit indices to identify meaningful dimensions of school climate. Given the 11 theoretical factors of school climate, we compared solutions with up to 11 factors. Because some of the items were dichotomous, we used tetrachoric correlations and the MLR estimator through MPlus (Muthén & Muthén, 2012). This method converts all variables to binary data and then computes correlations among variables based on rater agreement on each variable, with the assumption that the binary data represent underlying continuous variables (e.g. the extent to which different students feel safe or how much teasing they hear in one day). Thus, this method allows for the comparison of dichotomous and continuous data within one analysis (Muthén, 1984). A geomin rotation was used to improve the interpretability of factors. A geonim rotation is a type of oblique rotation that produces factor loadings and factor correlations similar to a confirmatory factor analysis, and is therefore useful for analyses in which there is expected to be overlap among factors (Hattori et al., 2017). Because school climate theoretically depends on both the school and the individual (Cornell et al., 2016; Gage et al., 2016), we compared standard deviations for each item across individuals within each school.

Table 2. Demographic information for schools in study 1.

	School 1	School 2	School 3	School 4	School 5	School 6	School 7	School 8	School 9
School Size	1606	1828	596	2261	647	225	385	2716	331
Locale	Urban	Suburban	Urban	Suburban	Suburban	Urban	Urban	Suburban	Rural
School Type	Public	Public	Magnet	Public	Public	Magnet	Charter	Public	Private
Free/reduced price lunch	83%	1%	36%	56%	16%	84%	Not reported	41%	Not reported
% Male	10%	52%	63%	55%	46%	68%	50%	51%	Not reported
Grades served	9–12	9–12	6–12	9–12	9–12	9–12	K-8*	9–12	9–12
Race/Ethnicity									
African American or Black	30%	2%	38%	7%	1%	78%	73%	17%	6%
American Indian or Alaska Native	1%	0%	0%	<1%	0%	1%	3%	<1%	0%
Asian or Asian-American	4%	6%	7%	4%	1%	3%	1%	7%	2%
Biracial or Multiracial	1%	3%	3%	7%	2%	1%	<1%	4%	3%
Hispanic or Latin	57%	6%	21%	40%	2%	14%	22%	36%	1%
Native Hawaiian or Pacific Islander	1%	0%	0%	1%	<1%	2%	0%	<1%	<1%
White or Caucasian	5%	83%	32%	41%	94%	<1%	1%	37%	88%

Note. School statistics were gathered from the National Center for Educational Statistics (NCES) database.
*Data was collected with grades 6–8.

Results

The results of the factor analysis indicated that the model with the best fit was a 10-factor solution, RMSEA = .028, SRMR = .034, CFI = .876, TLI = .840, χ^2 (3570) = 12,663.262, $p > .05$ ($\chi^2/743 = 3.547$). Although this model was the best fit and satisfied some of the standard criteria for goodness-of-fit (RMSEA < .08, SRMR < .08, χ^2 $p > .05$), it did not satisfy other criteria (CFI >.90, TLI >.95; e.g. Bentler, 1990; Steiger & Lind, 1980). Furthermore, factor loadings were relatively low (Kyriazos, 2018) and did not reveal theoretically-meaningful dimensions of school climate. For example, items with the highest loadings for the first factor included, 'The lunchroom was clean' (.47), 'Students were competitive' (.34), and 'Students' parents are involved at our school' (.28). Thus, none of the models examined were an adequate fit for the data.

Interpersonal variances for each item differed across schools (standard deviations across items ranged 0–2.0). For example, when asked whether they saw trash in the hallway, students from one school were about evenly divided (*SD* = .51), whereas students from another school answered unanimously (*SD* = 0). Thus, Items with the highest variance also differed by school. In one school, the highest variance emerged for the item, 'Most of the food available to us was nutritious' (*SD* = .51), whereas in another, it was, 'Students have a voice when school rules and regulations are created or changed' (*SD* = .51).

Discussion

This step revealed several important findings. First, the items did not represent meaningful factors. Upon reflection, this makes sense given that some of the items were more general (for example, asking about how safe one feels at school) whereas others were about specific, observable facets of the school, which may not necessarily correlate with each other or load onto latent constructs. For example, bullying rates and the availability of gender-neutral bathrooms could both contribute to a students' sense of psychological safety, but might not correlate with each other across schools. Based on these findings, we concluded that it was more appropriate to separate observable indicators from general items that were expected to converge into factors.

A second finding was that the item variances differed across schools. In other words, items on which students disagreed were not the same from school to school. This demonstrates the items contribute to discriminant validity and we therefore did not remove any items for failing to meet this requirement. This finding further supported our hypothesis that there are likely systematic differences in how students answer items; the ranges of scores (i.e. disagreements) were not random, but due to differences in school experience by demographic group. Next, items were presented via a web-based application so we could obtain feedback on the usability of the app.

Study 2: student feedback on walkthrough version 2

The purpose of this study was to obtain student feedback on the digital iteration of the Walkthrough. Specifically, we sought to determine how to improve survey administration and the presentation of results.

Figure 1. School climate walkthrough web-based app beta version. Sample item in which students swipe left to disagree or swipe right to agree.

Method

Students were recruited from two Connecticut public high schools (Schools E and F). Both schools were existing partners of the Yale Center for Emotional Intelligence through a youth empowerment programme called inspirED. After providing assent, students reviewed sample reports and provided feedback. Students then completed items via the application throughout the following school day. Researchers returned two days later for an additional student focus group covering student experiences using the app and viewing the report.

Participants

The first focus group at school E included 19 students (9 males, 10 females). The second focus group at school F included 9 students (2 male, 7 female). While additional demographic information was not collected from focus group participants, each school was instructed to invite a diverse group of students, representing all grades and ethnicities enrolled in the school. Table 1 provides additional information on the demographic and socioeconomic makeup of each school.

Materials

The Walkthrough was administered as a web-based application. Participants were presented with graphics representing school locations. As each location was selected, survey items appeared (e.g. 'I witnessed teasing in the bathrooms today'). Students responded by swiping left (no, not seen today), or right (yes, seen today). When a section was completed, the graphic for that location changed colour and students moved to the next section. See, Figure 1 for a screenshot of a survey item. The app did not yet have an automatic reporting feature, and thus pdf reports were provided back to students and school leaders.

[insert Figure 1 around here]

Results and discussion

In general, students reported favourable views of completing the survey through the application. They noted several areas for improvements, such as unclear instructions for the swiping feature. Students also recommended changing the survey timing; they explained that completing the survey during the day was distracting and that sometimes they witnessed things (e.g. teasing) after they already said 'No' to that item earlier in the day.

Students' suggested including a pinning feature to save items for later discussion, the option to filter results, and the capacity to easily share results with school leaders. They wanted reports to include raw numbers, percentages, and visual representations. Students were interested in seeing trends across the entire school as well as among groups, especially comparing their own answers to their peers'. Overall students' comments demonstrated that the survey and report sparked an interest in learning more about school climate and how to improve it (see, Table 3).

[insert Table 3 around here]

Students' feedback prompted several revisions. We replaced the colour scheme and added more options for viewing results. Students and educators wanted the tool to provide two types of information (composite scores and concrete feedback), yet one of the greatest concerns that emerged was the feasibility of answering questions throughout the day. Together with the Study 2 finding that the observable items did not form factors, we switched to a morning survey with a set of general items meant to produce traditional school climate scores and an afternoon checklist of observable items not meant to contribute to domain scores but simply to provide concrete feedback regarding potential improvements.

Study 3: refining items and factors for version 3

As mentioned in Study 1, the original items did not load onto meaningful factors, which meant that the assessment would not produce scores for each dimension of school climate, a crucial feature of utility. Therefore, we rewrote the items to more intentionally separate items that could be used to produce scores of school climate (general items) from those that offer a snapshot of how students experienced school that day (observable items). The first goal of this study was to identify the underlying factor structure for the revised general items. The second goal was to evaluate which observable items had the strongest correlations with the dimensions calculated from the general items.

Table 3. Student feedback on the web-based application.

Domain	Summary of Feedback	Sample Quotes
Use of technology	Positive: • Web-based is better than paper and pencil • Like the automatic and instantaneous scoring Negative: • Some were confused by swiping feature • Green/red colours for Yes/No were confusing	• *'Also because with our school climate surveys, like most of us don't see the results a lot and I think we're all interested in it too. So that would be really cool.'* • *'It reminded me of Trivia Crack how you have to press answers where you press answers, so then it was confusing.'*
Timing	Positive: • Felt shorter and more fun – more likely to complete Negative: • Issues accessing stable internet throughout the day • Difficulty remembering to complete survey • Silly to answer questions during day when might have to change answer later if something happens	• *'I felt like I wasn't doing what I was supposed to throughout the day.'* • *'. . . all the other questions I felt were sort of asking you . . . did this happen? But like I wouldn't know if that happened until I went through the entire day.'*
Report Organization	Suggestions: • Liked reports short, simple, and organised • Wanted nested results with clickable features • Options to filter and flag results • Wanted ability to share results with school leaders	• *'. . . you can click on the box and then that can bring you to in-depth information about that one specific section,'* • *'Maybe having, like, a summary of all the results for the school, and . . . a feature to email it to the leaders of the school.'*
Statistics	Suggestions: • Include the statistics for students to see • Include visuals with the numbers • Provide the raw numbers and percentages	• *'Definitely involve numbers. I feel numbers are a big thing; numbers don't lie'* • *'I always envision, like, a pie chart; you have colors, but you also have the numbers like right on them with like percentages'*
Student- and school-level results	Suggestions: • Want individualised results; reward for effort • Wanted to know overall averages; trends in the school	• *'Even like a . . . kind of like once you take it . . . it kind of gives feedback to us; like, you as a student and what your score is'* • *'. . . it needs to be like, "Okay overall this is the trend" because I think that's what a lot of people focus on'*
Group comparisons	Suggestions: • Show comparisons with other schools • Show comparisons among demographic groups within the school • Show individual results compared to school scores	• *'. . . addressing each side is important because you see a green bar . . . and you're like "OK it's not a problem" but there's still some people that it isn't.'* • *'. . . you can see if we're seeing the same problems so we can address a specific problem'*

Method

Three schools who had reached out to Yale Center for Emotional Intelligence for support with social-emotional learning expressed interest in testing this iteration of the assessment tool. Parent opt-out forms were sent home to all students enrolled at the three schools, and eligible students also completed an informed assent prior to participation. Participants completed the items on a single school day, using a link to the survey provided by the researchers and distributed by the schools. Participants were free to skip any question or stop participation at any time. General items were completed in the morning to help minimise the potential for students' responses to be biased by experiences from that specific day. Observable items, specific to that particular day, were completed at the end of the school day.

Participants

The sample included 759 students from three U.S. schools. This sample was large enough to satisfy recommendations for exploratory and confirmatory factor analyses (minimum $n > 300$; 5 participants *73 items = 365 participants; $n >$ number of parameters; DeVellis, 2017; Kyriazos, 2018), accounting for the sample being divided in half for factor analyses. Students' ages ranged from 14 to 18 ($M = 15.41$, $SD = 1.29$). Many students chose to skip the demographic questions: students identified as 34% female, 31% male, <1% another gender, 35% did not report. Students identified as Asian/Asian American (<1%), Black/African American (<1%), Biracial/Multiracial (6%), Hispanic/Latino (2%), Middle Eastern (<1%), White/Caucasian (55%), other (4%), and 33% who did not report their ethnicity. Participants identified as heterosexual (55%), bisexual (6%), gay or lesbian (2%), pansexual (1%), queer/other (1%), preferred not to answer (4%), and 33% did not report their sexual orientation. We have provided information on the demographic makeups of the three schools (Schools X, Y, and Z) as a whole in Table 4. Schools represented urban, suburban, and rural locales, private and public schools, a range of sizes and grades served, as well as ethnic composition.

[insert Table 4 about here]

Materials

School climate items were revised to more directly address two goals of the assessment tool. First, we extended the list of general items that asked about students' overall perceptions of their school's climate (e.g. 'Adults at my school are friendly towards each other'). These 73 items were rated on a 5-point Likert scale from 1 (*Strongly disagree*) to 5 (*Strongly agree*). Second, we extended the list of items that asked about specific, observable phenomena related to each dimension of school climate (e.g. 'Today I observed adults at my school being rude to each other'). Students indicated whether they observed each of these 117 items on the day of the assessment. Students could skip or select 'not applicable' to any item as needed. The number of general and observable items for each dimension were (respectively): adult relationships (6, 6), respect for

Table 4. Demographic information for schools in study 3.

	School X	School Y	School Z
School Size	401	2716	536
Locale	Urban	Suburban	Rural
School Type	Private	Public	Public
Free/reduced price lunch	Not reported	41%	99%
% Male	61%	51%	48%
Grades served	6–12	9–12	7–8
Race/Ethnicity			
African American or Black	2%	17%	35%
American Indian or Alaska Native	0%	<1%	<1%
Asian or Asian-American	3%	7%	1%
Biracial or Multiracial	3%	4%	4%
Hispanic or Latin	69%	36%	6%
Native Hawaiian or Pacific Islander	<1%	<1%	0%
White or Caucasian	23%	37%	54%

Note. School statistics were gathered from the National Center for Educational Statistics (NCES) database for schools Y and Z, and by self-report from school X.

diversity (5, 10), student-adult relationships (7, 9), student relationships (7, 12), emotional safety (6, 18), physical safety (5, 8), rules and norms (6, 8), connectedness (7, 6), physical space (5, 18), effective teaching practices (4, 11), and support for SEL (8, 7).

Data analyses

Students' responses were randomly divided into two datasets. We used the first dataset (n = 379) to conduct a principal axis factor analysis to identify the factor structure that best represented variance within the general items. We used an oblique rotation to improve the interpretability of factors due to theoretical overlap among factors. To determine the number of general items comprising each dimension, we considered the size of factor loadings as well as theory – specifically, that there were items that represented all aspects of that dimension. Items with loadings <.35 on any dimension were removed. With the second database (n = 380), we conducted a confirmatory factor analysis to evaluate the model fit identified in the first step. Analyses were performed in MPlus using MLR to estimate model parameters and goodness-of-fit (Muthen & Muthen, 2012). The specified model was the nine-factor solution identified in the exploratory factor analysis. Next, composite scores were computed based on the factors identified through the factor analyses (using the full dataset). We then conducted Pearson bivariate correlations for the observable items and each factor. For example, we computed the correlation between the endorsement of the statement, 'Today, I saw or heard about physical violence taking place in/on/around classrooms' with students' sense of safety. These correlations were calculated for each school and then averaged across schools. These were used to select items.

Results

From the original database (N = 1,086), we removed 328 respondents who failed an attention check (e.g. did not follow the instructions, 'Select 1 for this item') or typed nonsensical responses (e.g. 'helicopter' as their gender), leaving 759 cases. No cases were removed for missing data. For general items, skewness ranged from 0.85 to −1.15, with most items (92%) being negatively skewed. Given that the distributions were relatively

normal, no transformations were conducted. There were no apparent outliers or non-linear relationships among items. There were no significant differences between the two halves of the samples with regards to age, $t(1457) = 3.62$, $p = .72$; gender, $x^2(6) = 8.62$, $p = .20$; ethnicity, $x^2(6) = 4.18$, $p = .65$; or sexual orientation, $x^2(6) = 4.18$, $p = .65$. Data from the first half of the database satisfied the statistical assumptions for an exploratory factor analysis (KMO = .96, Bartlett's Test $p < .001$).

The goal of the exploratory factor analysis was to identify a factor structure that was parsimonious and accounted for the most variance possible. First, we explored a unidimensional factor solution, which explained 31.99% of the variance in scores (Eigenvalue = 23.35). Next, based on Kaiser (1960), we considered a solution that included all factors with Eigenvalues greater than 1.00. This 12-factor solution accounted for 60.12% of the cumulative variance but did not produce theoretically-meaningful dimensions. Next, we inspected the scree plot and factor loadings for each potential solution between 1 and 12 factors to determine which solution accounted for the most variance while also producing meaningful factor loadings. Based on this information, we selected a nine-factor solution, which accounted for 55% cumulative variance in scores. Factor correlations ranged from −.43 to .42. See, Table 5 for information about this solution. See, Table 6 for Pearson bivariate correlations among factors.

[insert Table 5 around here]

[insert Table 6 around here]

The results of the confirmatory factor analysis results indicated a good model fit for the nine-factor solution, RMSEA = .051, SRMR = .013, CFI = .987, TLI > 0.985, $\chi^2(743) = 21.87.572$ ($\chi^2/743 = 2.943$), $p < .05$. These results fulfill common standards for assessing goodness-of-fit (RMSEA < .08, SRMR < .08, CFI > .90, TLI > .95, χ^2 $p < .05$; Bentler, 1990; Steiger & Lind, 1980). Correlations between observable items and dimensions ranged from −.21 to .33. See Supplemental Table 1 for a full correlation matrix. We used these data to determine which items to retain for the final scale. First, we considered the 10 items that had the strongest correlation with each dimension; we then used the lowest correlation among these items, $r = |.125|$, as a cut-off for including other observable items. This cut-off is similar to the strength of correlations that have been found in previous studies comparing ratings of the same construct either globally or across a day, such as daily/global loneliness ($r = .27$; Doane & Adam, 2010), and global meaning with daily behaviours expected to enhance meaning ($r = .12$ to .22; Steger & Kashdan, 2007). Through this process, 46 items were removed. The remaining 73 observable items were each correlated with at least one dimension, and most (69) were correlated with more

Table 5. Walkthrough measure – final factor structure.

Factors	Name	Eigenvalues	Number of items selected	Factor loading ranges
1	Supportive Teaching Practices	22.6	7	.46-.68
2	Physical Safety	3.49	6	.54-.75
3	Adult Relationships	3.01	4	.74-.79
4	Emotional Safety	2.26	2	.69-.78
5	Student Relationships	2.05	5	.35-.58
6	Student-adult Relationships	2.02	5	.52-.68
7	School Pride	1.76	6	.57-.77
8	Respect for Diversity	1.52	5	.59-.76
9	Social Safety	1.47	2	.70-.74

Table 6. Correlations among dimensions of school climate, study 2.

	1	2	3	4	5	6	7	8	9
1. Physical Safety	.90								
2. Social Safety	.15*	.87							
3. Emotional Safety	.40**	.17*	.88						
4. Respect-Diversity	.53**	.11*	.42**	.83					
5. Student-Adult Rel.	.54**	.09*	.49**	.63**	.85				
6. Student Rel.	.48**	.19*	.51**	.55**	.57**	.87			
7. Adult Rel.	.49**	.07	.35**	.53**	.58**	.48**	.89		
8. Teaching Practices	.55**	.13*	.46**	.63**	.71**	.59**	.58**	.76	
9. School Pride	.48**	.11*	.45**	.55**	.57**	.59**	.49**	.65**	.75

Notes. Rel. = Relationships. Alpha values are listed in the diagonal. A Bonferroni correction was applied. *$p < .01$, **$p < .001$

than one dimension. Notably, of several items that correlated with all 9 dimensions, one was *'Today my teachers paid attention to all students'*. Each dimension was represented by 10–58 observable items.

The final assessment tool included 42 general items reflecting nine dimensions of school climate ($\alpha = .75$-$.91$) and 73 observable items that were correlated with those nine dimensions ($r = |.125|$-$|.33|$).

Discussion

The primary goal of this study was to determine the factor structure of the school climate items. Results from exploratory and confirmatory factor analyses support a nine-factor solution, representing nine dimensions of school climate: physical safety, emotional safety, social safety, relationships among students, relationships among adults, relationships between students and adults, supportive teaching practices, support for diversity, and school pride. These dimensions align closely with those of other prominent school climate assessment tools, including the Comprehensive School Climate Inventory (National School Climate Council, 2007) and the U.S. Department of Education School Climate Survey (EDSCLS, 2016). Notably, students' ratings of effective teaching practices and teacher support for social and emotional learning loaded on the same dimensions, indicating that, in students' views, teaching quality includes both academic instruction and care towards students' overall well-being. Additionally, the item indicating that teachers paid attention to all students, which correlated with all nine dimensions, aligns with previous research, which shows that whether students have a trusted adult at school who cares about them is one of the most meaningful predictors of adolescent health and well-being (Murphey et al., 2013; Pisani et al., 2013; Sieving et al., 2017).

A second finding was that observable items correlated with the general dimensions, in support of the idea that daily observations relate to overall perceptions of school climate. Small to moderate correlation sizes were expected given that each observable item reflects experience from a particular day, whereas general dimensions reflect students' broader perceptions. Zullig et al. (2010) discuss this tension between items where participants use their own criteria (e.g. items such as 'My school is safe'), and items in which the criteria is decided for the participant and may not be how they would have defined it for themselves (e.g. 'The front office requires visitors to check in'). Moreover, certain experiences can influence how students view several facets of their school's climate as we saw with

observable items that were related to multiple dimensions of school climate. For example, negative interactions with security guards was related to both student-adult relationships and respect for diversity. Thus, we follow the conclusion of Zullig et al. (2010) that a full picture of a students' school experience is best drawn from a combination of surveys gathering subjects' general opinions–which reflect perceptions of school climate as latent constructs, and are more likely to be consistent across time–and observable items that may not be as strongly tied to a latent construct, and that could fluctuate from day to day, but that also could offer useful insights into the specific facets of students' experiences that could be changed to improve students' perceptions of school climate.

This study has an important limitation: although dividing a sample in half is a common practice for cross-validating factor structures (e.g. Brown, 2015; Kyriazos, 2018; Thompson, 2013), this procedure does not account for sample idiosyncrasies. Therefore, a useful next step would be to replicate the factor structure in a separate sample to confirm that it generalises across other schools and students.

General discussion

The purpose of this paper was to illustrate the process of building an educational assessment tool that maximises utility for students: one that is engaging, meaningful, and inclusive. This process was applied to the development of the *School Climate Walkthrough Tool* (see, Table 7 for a detailed description). A critical feature of this process was the collection of feedback from experts in social and emotional learning, educators, and students, and a true embracing of the historic slogan *'Nothing about us, without us'* (Charlton, 1998). Each round of feedback yielded valuable insights that shaped the final tool.

[insert Table 7 around here]

A second key element of this process was administering the assessment through an online application. Digitally-administered assessments eliminate printing and shipping costs as well as potential errors in scoring. Additionally, applications that automatically provide reports decrease turnaround time and reduce or eliminate scoring fees. This is especially helpful for constructs such as school climate where the results lose relevance as students advance and graduate. Technology also allows for interactive features such as toggling between total scores and group breakdowns (e.g. scores by grade level). Figure 2 provides an example of how group breakdowns are presented within the School Climate Walkthrough App. Moreover, students reported that an aesthetically-pleasing and interactive application was more engaging than typical paper-and-pencil or online surveys.

A third critical step for understanding school climate is to explore group differences in how students' experience school. For example, a large study of high school students found that, after nesting students in schools and controlling for other student characteristics (e.g. gender, parent education), perceptions of school climate differed between Black and White students (Konold et al., 2017). Studies using person-centred approaches (e.g. latent class analysis) have also found within-school heterogeneity of student perceptions of climate along racial lines (Shukla et al., 2016). The Walkthrough facilitates group comparisons by asking questions such as, 'I have an educator I can go to for support' and then displaying systematic differences in how students' feelings of support vary across groups (e.g. grade, gender, ethnicity, sexual orientation). This method asks students to

Table 7. School climate walkthrough measure.

Dimension	General Item Examples	Observable Item Examples
Physical safety- Do students feel physically safe at school?	• My school has clear rules about students' behaviour. • I feel safe from physical harm at my school.	• Today rules were enforced fairly in/around the lunchroom. • Today, I saw or heard about physical violence taking place in/on/around bathrooms.
Emotional safety- Do students feel safe expressing their emotions at school?	• I feel comfortable expressing my emotions at school. • I feel comfortable talking about my interests in school.	• I felt comfortable expressing my opinions during class today. • Today, we talked about emotions at school.
Social safety – Do students feel safe from teasing and exclusion at school?	• I worry about getting teased at school. • I worry that I will get left out of things at school.	• Today, I heard unkind or offensive language in/on/around the hallway, stairs, and/or elevators. • Today I heard students gossiping or spreading rumours about other students.
Student relationships- Do students have friendly, supportive relationships?	• Students at my school get along with each other. • Students at my school support each other emotionally.	• Today I saw a student go out of his or her way to help another student. • Today I saw students ignore each other or cut each other off in class.
Adult relationships- Do adults at the school have friendly, supportive relationships?	• Adults at my school are friendly towards each other. • Teachers at my school support each other.	• Today, I observed adults at my school being rude towards each other. • When teachers passed each other in the hallways, they greeted each other in a friendly way.
Student-adult relationships- Do students and adults at the school have caring, supportive relationships with each other?	• Teachers at my school genuinely care about students. • I have an educator that I can go to for support.	• Today, I saw adults greeting students on their way into school. • Today, I saw teachers making an effort to build positive relationships with their students during class.
Supportive teaching practices- Do teachers use effective strategies for supporting academic learning and SEL?	• I feel like I learn a lot in my classes. • My school helps students learn how to better understand and regulate their emotions.	• Today, we talked about emotions at school. • Today, my teachers provided challenges to help students grow.
Support for diversity- Are students of diverse backgrounds, abilities, and experiences respected and supported?	• My school is accepting of people from all backgrounds. • My school offers opportunities to learn about diverse perspectives.	• Today I saw someone in my school being discriminated against based on race/ethnicity. • We talked about issues related to equity, inclusion, and/or diversity at school today.
School pride- Do students take pride in their school community?	• I like my school's overall appearance. • I am proud to say that I go to my school.	• Today, I saw a mess in/around entrances. • Today I heard an adult talk about my school as a community.

only comment on their own experience, which is essential for some items. For example, students would not be able to comment on whether all bathrooms are equitably supplied as most students frequent only the bathrooms aligned with their gender identity. However, asking, 'There were enough supplies in the bathrooms today' does allow for such comparisons. In focus groups, students were particularly interested in these group differences, indicating that these data may help to spark additional conversations about students' experiences at school and how to support all students. Assessing how students'

Figure 2. Examples of demographic discrepancies provided by school climate walkthrough reports. This figure shows a discrepancy by gender identity detected on the social safety dimension and an associated discrepancy from an observable item.

experiences of school vary can help to identify inequitable, exclusionary, and/or harmful practices, and therefore is a critical first step for schools to reduce gaps in achievement, engagement, and exclusionary discipline practices.

A final set of learnings was on generating useful feedback for students and school leaders. We maximised the utility of feedback by including two types of items: (1) general items reflecting overall perceptions of school climate which could be used to draw comparisons across and within schools; and (2) observable items which provided clear direction for next steps. The two item types worked together; for example, if a school received a low score on general perceptions of physical safety, they could review the observable data to identify where students saw physical violence. For both types of items, we highlighted subgroup differences within their reports (see, Figure 2). Students' feedback revealed that they were especially interested in using group differences as launching points for school climate improvement projects.

Through sharing back data with students in focus groups, we witnessed how information motivates students to improve their school. Consequently, the Walkthrough could be used at the beginning of programming such as that provided by inspirED – a free resource that empowers students to form teams and enhance their school climate through projects and campaigns (Hoffmann et al., n.d.; inspiredstudents.org) – to provide students with direction and motivation to improve their school climate. Given the potential value of sharing results with students, we recommend that educational assessment tools include reports designed specifically for students, and that these are shared as soon as possible.

Limitations and future directions

We note that there are instances in which some of the steps of the process we present might not be feasible or appropriate. For example, web-based tools may not be possible in areas with limited internet connectivity. In such cases, assessment developers can offer paper-based forms, with scoring software or services (such as those described in Study 1), or allow students to complete the assessment offline and upload responses later. Furthermore, questions around sensitive topics (e.g. depression, abuse) that may yield

responses in which mandated reporters would need to take action should not be applied to this format of anonymous responding (precluding action), nor to automated scoring procedures that are openly shared. Researchers must evaluate whether each feature and process is appropriate for their construct and the context in which the tool will be used. Moreover, we note the continuous need to scrutinise and adjust the terminology used, demographic categories analysed, and school practices assessed to remain current and meaningful alongside the evolution of human science.

An essential next step, after developing an assessment tool, is to test its psychometric properties. Thus, the Walkthrough should be subjected to tests of convergent validity, test-retest reliability, sensitivity to change over time or as a result of intervention, incremental validity in comparison to other school climate assessment tools, discriminant validity with related constructs such as personality, and predictive validity of student and school outcomes. This work should include representative samples, which is essential for ensuring that this tool is valid for all students and school contexts, and contribute to the uncovering and resolution of discrepancies in students' school experiences. Not collecting student demographic information in our early studies was a limitation. While not perfect, we hope that the process described here illustrates the many benefits of including student voice from the very first steps of developing an assessment tool, increasing the likelihood of a meaningful, useful, and engaging final product that is used not just to rate schools, but to transform them.

Acknowledgments

This work was supported by a grant from the Chan Zuckerberg Initiative DAF, an advised fund of the Silicon Valley Community Foundation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Chan Zuckerberg Initiative [GR101828].

Notes on contributors

Jessica D. Hoffmann, Ph.D., is a research scientist at the Yale Center for Emotional Intelligence. She serves as the director of RULER for secondary schools, which helps bring social-emotional learning to middle and high school students and their educators to promote safe, nurturing environments where students and educators can reach their full potential. Jessica received her B. A. in psychology and sociology from Brandeis University, and her M.A. and Ph.D. degrees in clinical psychology from Case Western Reserve University. Jessica specialises in working within school settings to develop creative interventions aimed at enhancing children's creative thinking to promote adaptive coping and enhance mental health. Her current research focuses on the impact of emotion skills instruction on school climate, creative problem solving, and emotion regulation ability.

Rachel Baumsteiger, Ph.D., is an associate research scientist at the Yale Center for Emotional Intelligence. Her research focuses on moral development and well-being among adolescents and emerging adults. Her most recent projects involved developing strategies for assessing prosocial behaviour and developing tools for fostering gratitude, purpose, and prosocial habits among adolescents and young adults. Rachel is also interested in investigating instructional strategies for enhancing student well-being in post-secondary settings. Rachel joined the Yale Center for Emotional Intelligence after receiving an M.A. and Ph.D. in psychology from the Claremont Graduate University with a concentration in positive developmental psychology, and a B.A. in psychology from the University of South Florida. Her work at the Center focuses on developing methods for assessing and promoting emotional intelligence and school climate among high school students.

Jennifer Seibyl, M.Ed., is a research associate at the Yale Center for Emotional Intelligence. She works on RULER for secondary schools, inspirED, and additional research projects. Jenny's initial introduction into the world of emotional intelligence began in the Summer of 2013 when she was an intern for the Yale Center for Emotional Intelligence, working on the nascent stages of preschool RULER and the RELATE project. She continued this thread of interest throughout her academic career. Jenny's most recent work experience was at McLean Hospital on a male adolescent residential unit treating young men with emerging borderline personality traits. This unique subset of clients sparked her interest in the intersection between youth mental health outcomes and SEL. Jenny holds a B.A. in Psychology from Skidmore College and a M.Ed. in Human Development and Psychology from Harvard University.

Elinor Hills, M.Sc., is a Brooklyn-born photographer who graduated from Yale University in 2018 with a degree in art and anthropology. Elinor worked as a postgraduate associate at the Yale Center for Emotional Intelligence for two years before pursuing a Master of Science degree in Medical Anthropology at Oxford University. Across the disciplines of anthropology, photography, and design, her research interests include kinship, child development, family mental health & emotional wellbeing, language and communication in healthcare settings, healthcare equity/system reform, and the importance of visual narratives and communication in making medicine more accessible.

Christina Bradley is a doctoral student in the Management & Organizations department at the Ross School of Business. She studies how emotions affect the way we think, feel, and behave in situations at work. She hopes her research will help individuals and organisations become curious about and skilled with handling emotions, primarily in relational contexts. Christina believes that understanding and developing the various skills of emotional intelligence can help individuals both perform better at work and live healthier, more meaningful lives. Prior to joining the Management & Organizations PhD program at Michigan Ross, Christina completed her undergraduate degree in psychology at Yale University. During and after graduation, she conducted research on emotions and emotional intelligence with both the Yale Center for Emotional Intelligence and the Contentment Foundation.

Christina Cipriano, Ph.D., holds a dual appointment at the Yale School of Medicine as Assistant Professor of Applied Developmental and Educational Psychology at the Yale Child Study Center and Director of Research at the Yale Center for Emotional Intelligence (YCEI). Dr. Cipriano's research focuses on the systematic examination of social and emotional learning (SEL) in the promotion of pathways to optimal developmental outcomes for traditionally marginalised student and teacher populations. Dr. Cipriano received her Ph.D. in Applied Developmental and Educational Psychology from Boston College, her Ed.M. from the Harvard Graduate School of Education and undergraduate degree from Hofstra University. Dr. Cipriano has extensive experience working in classrooms and schools, providing training to teachers and support staff, and direct instruction to students. As the Director of Research at the YCEI, Dr. Cipriano oversees, designs, and maintains large-scale basic and applied research studies on: emotions, emotional intelligence, and SEL, including longitudinal studies and randomised controlled trials.

Dr. Cipriano has a wealth of project management experience as the PI and Director of numerous major grants supporting the development and validation of SEL assessments, classroom observation practices, and SEL evaluation and efficacy trials.

Marc A. Brackett, Ph.D., is founding director of the Yale Center for Emotional Intelligence and professor in the Child Study Center, Yale School of Medicine. His grant-funded research focuses on: (1) the role of emotions and emotional intelligence in learning, decision making, creativity, relationship quality, and mental health; (2) the measurement of emotional intelligence; and (3) the influences of emotional intelligence training on children's and adults' health, performance, and workplace performance and climate. Marc has published 125 scholarly articles and has received numerous awards, including the Joseph E. Zins Award and an honorary doctorate from Manhattanville College. He also is a distinguished scientist on the National Commission on Social, Emotional, and Academic Development and on the board of directors for the Collaborative for Academic Social, and Emotional Learning (CASEL).

Marc is the lead developer of RULER, a systemic, evidence-based approach to social and emotional learning that has been adopted by over 2,000 schools across the United States and other countries. Marc is the author of *Permission to Feel* (Celadon/Macmillan), which was released in September of 2019. Marc regularly consults with large companies on best practices for integrating the principles of emotional intelligence into training and product design. He is co-founder of Oji Life Lab, a corporate learning firm that develops innovative digital learning systems for emotional intelligence.

ORCID

Jessica D. Hoffmann http://orcid.org/0000-0003-4339-2836
Rachel Baumsteiger http://orcid.org/0000-0002-1506-3907
Christina Cipriano http://orcid.org/0000-0002-7414-1821
Marc A. Brackett http://orcid.org/0000-0002-5245-0244

References

Assessment Work Group. (2019). *Student social and emotional competence assessment: the current state of the field and a vision for its future*. Collaborative for Academic, Social, and Emotional Learning.

Bandyopadhyay, S., Cornell, D. G., & Konold, T. R. (2009). Validity of three school climate scales to assess bullying, aggressive attitudes, and help seeking. *School Psychology Review*, *38*(3), 338–355. https://doi.org/10.1080/02796015.2009.12087819

Bear, G. G., Gaskins, C., Blank, J., & Chen, F. F. (2011). Delaware school climate survey-student: its factor structure, concurrent validity, and reliability. *Journal of School Psychology*, *49*(2), 157–174. https://doi.org/10.1016/j.jsp.2011.01.001

Bennett, R. E. (2018). Educational assessment: what to watch in a rapidly changing world. *Educational Measurement: Issues and Practice*, *37*(4), 7–15. https://doi.org/10.1111/emip.12231

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Berry, D. T. R., Wetter, M. W., Baer, R. A., Larsen, L., Clark, C., & Monroe, K. (1992). MMPI-2 random responding indices: validation using self-report methodology. *Psychological Assessment*, *4*(3), 340–345. https://doi.org/10.1037/1040-3590.4.3.340

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Bottiani, J. H., Bradshaw, C. P., & Mendelson, T. (2017). A multilevel examination of racial disparities in high school discipline: black and white adolescents' perceived equity, school belonging, and adjustment problems. *Journal of Educational Psychology*, *109*(4), 532–545. https://doi.org/10.1037/edu0000155

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Publications.

Charlton, J. I. (1998). *Nothing about us, without us: disability, oppression and empowerment*. University of California Press.

Clawar, M., Thompson, K. W., & Pink, G. (2018). *Range matters: rural averages can conceal important information*. North Carolina Rural Health Research Program.

Cohen, J., McCabe, E. M., Michelli, N., & Pickeral, T. (2009). School climate: research, policy, practice, and teacher education. *Teachers College Records*, *111*(1), 180–213. https://doi.org/10.1177/016146810911100108

Cornell University. (2012) . *National Social Survey*. Cornell Publishing.

Cornell, D., Klein, J., Konold, T., & Huang, F. (2012). Effects of validity screening items on adolescent survey data. *Psychological Assessment*, *24*(1), 21–35. https://doi.org/10.1037/a0024824

Cornell, D. G., & Loper, A. B. (1998). Assessment of violence and other high-risk behaviors with a school survey. *School Psychology Review*, *27*(2), 317–330. https://doi.org/10.1080/02796015.1998.12085918

Cornell, D., Shukla, K., & Konold, T. (2016). Authoritative school climate and student academic engagement, grades, and aspirations in middle and high schools. *AERA Open*, *2*(2), 1–18. https://doi.org/10.1177/2332858416633184

Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, *10*(1), 3–31. https://doi.org/10.1177/001316445001000101

DeRosier, M. E., & Thomas, J. M. (2018). Establishing the criterion validity of zoo U's game-based social emotional skills assessment for school-based outcomes. *Journal of Applied Developmental Psychology*, *55*, 52–61. https://doi.org/10.1016/j.appdev.2017.03.001

Detert, J. R., & Burris, E. R. H. R. B. (2016). *Can your employees really speak freely?*. *Harvard Business Review*, *94*(1), 80–87. https://hbr.org/2016/01/can-your-employees-really-speak-freely

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage.

Doane, L. D., & Adam, E. K. (2010). Loneliness and cortisol: momentary, day-to-day, and trait associations. *Psychoneuroendocrinology*, *35*(3), 430–441. https://doi.org/10.1016/j.psyneuen.2009.08.005

Dubad, M., Winsper, C., Meyer, C., Livanou, M., & Marwaha, S. (2018). A systematic review of the psychometric properties, usability and clinical impacts of mobile mood-monitoring applications in young people. *Psychological Medicine*, *48*(2), 208–228. https://doi.org/10.1017/s0033291717001659

Family Foundation, K. (2010). *Generation M2: media in the lives of 8-18 year-olds*.

Fatou, N., & Kubiszewski, V. (2018). Are perceived school climate dimensions predictive of students' engagement? *Social Psychology of Education: An International Journal*, *21*(2), 427–446. https://doi.org/10.1007/s11218-017-9422-x

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, *18*(9), 27–32. https://doi.org/10.3102/0013189x018009027

Furlong, M. J., Greif, J. L., Bates, M. P., Whipple, A. D., Jimenez, T. C., & Morrison, R. (2005). Development of the California school climate and safety survey-short form. *Psychology in the Schools*, *42*(2), 137–149. https://doi.org/10.1037/t77531-000

Furlong, M., Sharkey, J., Bates, M. P., & Smith, D. (2004). An examination of reliability, data screening procedures, and extreme response patterns for the youth risk behavior surveillance survey. *Journal of School Violence*, *3*(2–3), 109–130. https://doi.org/10.1300/J202v03n02_07

Gage, N. A., Larson, A., Sugai, G., & Chafouleas, S. M. (2016). Student perceptions of school climate as predictors of office discipline referrals. *American Educational Research Journal*, *53*(3), 492–515. https://doi.org/10.3102/0002831216637349

Grazia, V., & Molinari, L. (2021). School climate multidimensionality and measurement: A systematic literature review. *Research Papers in Education*, *36*(5), 561–587. https://doi.org/10.1080/02671522.2019.1697735

Griffith, J. (1997). Student and parent perceptions of school social environment: Are they group based? *The Elementary School Journal*, *98*(2), 135–150. https://doi.org/10.1086/461888

Hattori, M., Zhang, G., & Preacher, K. J. (2017). Multiple local solutions and geomin rotation. *Multivariate Behavioral Research*, *52*(6), 720–731. https://doi.org/10.1080/00273171.2017.1361312

Hoffmann, J. D., McGarry, J. A., Baumsteiger, R., Seibyl, J., & Brackett, M. A. (n.d.). Emotional empowerment in high school life. In G. Misra & I. Misra (Eds.), *Emotions in cultural context*. Springer.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Konold, T., Cornell, D., Shukla, K., & Huang, F. (2017). Racial/ethnic differences in perceptions of school climate and its association with student engagement and peer aggression. *Journal of Youth and Adolescence*, *46*(6), 1289–1303. https://doi.org/10.1007/s10964-016-0576-1

Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In J. D. Wright & P. V. Marsden (Eds.), *Handbook of survey research* (Second ed; pp. 263–314). Emerald Group.

Kyriazos, T. A. (2018). Applied psychometrics: the 3-faced construct validation method, a routine for evaluating a factor structure. *Psychology*, *9*(8), 2044–2072. https://doi.org/10.4236/psych.2018.98117

Lenz, A., Rocha, L., & Aras, Y. (2021). Measuring school Climate: A systematic review of initial development and validation studies. *International Journal for the Advancement of Counseling*, *43*(1), 1–15. https://doi.org/10.1007/s10447-020-09415-9

Loukas, A., Suzuki, R., & Horton, K. D. (2006). Examining school connectedness as a mediator of school climate effects. *Journal of Research on Adolescence*, *16*(3), 491–502. https://doi.org/10.1111/j.1532-7795.2006.00504.x

Marraccini, M. E., Fang, Y., Levine, S. P., Chin, A. J., & Pittleman, C. (2020). Measuring student perceptions of school climate: A systematic review and ecological content analysis. *School Mental Health*, *12*(2), 195–221. https://doi.org/10.1007/s12310-019-09348-8

Mavletova, A. (2014). Web surveys among children and adolescents: is there a gamification effect? *Social Science Computer Review*, *33*(3), 372–398. https://doi.org/10.1177/0894439314545316

McClain, L., Gulbis, A., & Hays, D. (2018). Honesty on student evaluations of teaching: Effectiveness, purpose, and timing matter! *Assessment & Evaluation in Higher Education*, *43*(3), 369–385. https://doi.org/10.1080/02602938.2017.1350828

Meier, S. T. (1994). *The chronic crisis in psychological measurement and assessment: A historical survey*. Academic Press.

Murphey, D., Bandy, T., Schmitz, H., & Moore, K. A. (2013). *Caring adults: important for positive child well-being*. Child Trends.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*(1), 115–132. https://doi.org/10.1007/bf02294210

Muthen, L. K., & Muthen, B. (2012). *Mplus user's guide* 7th(ed.). 1998-2012. Muthen & Muthen.

National Center for Education Statistics. (2016). *Education Department School Climate Survey (EDSCLS)*.https://safesupportivelearning.ed.gov/edscls/administration

National School Climate Council. (2007). *The school climate challenge: narrowing the gap between school climate research and school climate policy, practice guidelines and teacher education policy*. http://www.schoolclimate.org/climate/advocacy.php

Olsen, J., Preston, A. I., Algozzine, B., Algozzine, K., & Cusumano, D. (2018). A review and analysis of selected school climate measures. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *91*(2), 47–58. https://doi.org/10.1080/00098655.2017.1385999

Osborne, J. W., & Blanchard, M. R. (2011). Random responding from participants is a threat to the validity of social science research results. *Frontiers in Psychology*, *1*, 1–7. https://doi.org/10.3389/fpsyg.2010.00220

Pellegrino, J. W. (2020). Important considerations for assessment to function in the service of education. *Educational Measurement: Issues and Practice*, *39*(3), 81–85. https://doi.org/10.1111/emip.12372

Pisani, A. R., Wyman, P. A., Petrova, M., Schmeelk-Cone, K., Goldston, D. B., Xia, Y., & Gould, M. S. (2013). Emotion regulation difficulties, youth-adult relationships, and suicide attempts among high school students in underserved communities. *Journal of Youth and Adolescence*, *42*(6), 807–820. https://doi.org/10.1007/s10964-012-9884-2

Ramsey, C. M., Spira, A. P., Parisi, J. M., & Rebok, G. W. (2016). School climate: perceptual differences between students, parents, and school staff. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, *27*(4), 629–641. https://doi.org/10.1080/09243453.2016.1199436

Revilla, M., & Ochoa, C. (2017). Ideal and maximum length for a web survey. *International Journal of Market Research*, *59*(5), 557–565 http://doi.org/10.2501/IJMR-2017-039 .

Rideout, V., & Robb, M. B. (2019). *The common sense census: media use by tweens and teens, 2019*. Common Sense Media.

Romero, L. S., & O'Malley, M. D. (2020). An examination of classes of school climate perceptions among Latinx middle school students. *Journal of School Psychology*, *82*(October), 70–84. https://doi.org/10.1016/j.jsp.2020.08.001

Rust, J., & Golombok, S. (2014). *Modern psychometrics: the science of psychological assessment* (*Third Edition.* ed.). Routledge.

Shukla, K., Konold, T., & Cornell, D. (2016). Profiles of student perceptions of school climate: relations with risk behaviors and academic outcomes. *American Journal of Community Psychology*, *57*(3–4), 291–307. https://doi.org/10.1002/ajcp.12044

Sieving, R. E., McRee, A. L., McMorris, B. J., Shlafer, R. J., Gower, A. L., Kapa, H. M., Resnick, M. D., Doty, J. L., Plowman, S. L., & Resnick, M. D. (2017). Youth-adult connectedness: A key protective factor for adolescent health. *American Journal of Preventive Medicine*, *52*(3), S275–S278. https://doi.org/10.1016/j.amepre.2016.07.037

Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology*, *55*(1), 167–194. https://doi.org/10.1111/j.1744-6570.2002.tb00108.x

Steger, M. F., & Kashdan, T. B. (2007). Stability and specificity of meaning in life and life satisfaction over one year. *Journal of Happiness Studies*, *8*(2), 161–179. https://doi.org/10.1007/s10902-006-9011.8

Steiger, J. H., & Lind, J. C. (1980, May). Statistically-based tests for the number of common factors. *Paper presented at the annual spring meeting of the psychometric society*, Iowa City.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, *83*(3), 357–385. https://doi.org/10.3102/0034654313483907

Thompson, B. (2013). Overview of traditional/classical statistical approaches. In T. Little (Ed.), *The Oxford handbook of quantitative methods* (Vols. 1-2, pp. 7–25). Oxford University Press.

U.S. Department of Education. (2020). *Characteristics of children's families. Condition of Education*. https://nces.ed.gov/programs/coe/pdf/coe_cce.pdf

U.S. Education Council. (2019). *Measuring school climate and social and emotional development: A navigation guide for states and districts*. https://ccsso.org/sites/default/files/2019-03/CCSSO.EdCounsel%20SE%20and%20School%20Climate%20measurement_0.pdf

Van Eck, K., Johnson, S. R., Bettencourt, A., & Johnson, S. L. (2017). How school climate relates to chronic absence: A multi-level latent profile analysis. *Journal of School Psychology*, *61* (April), 89–102. https://doi.org/10.1016/j.jsp.2016.10.001 .

Varjas, K., Henrich, C. C., & Meyers, J. (2009). Urban middle school students' perceptions of bullying, cyberbullying, and school safety. *Journal of School Violence*, *8*(2), 159–176. https://doi.org/10.1080/15388220802074165

Wang, M.-T., & Degol, J. L. (2016). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*, *28*(2), 315–352. https://doi.org/10.1007/s10648-015-9319-1

Wang, M., & Dishion, T. J. (2012). The trajectories of adolescents' perceptions of school climate, deviant peer affiliation, and behavioral problems during the middle school years. *Journal of Research on Adolescence*, *22*(1), 40–53. https://doi.org/10.1111/j.1532-7795.2011.00763.x

Wise, S. L. (2006). An investigation of the differential effort received by items on a low-stakes computer-based test. *Applied Measurement in Education*, *19*(2), 95–114. https://doi.org/10.1207/s15324818ame1902_2

Zullig, K. J., Koopman, T. M., Patton, J. M., & Ubbes, V. A. (2010). School climate: historical review, instrument development, and school assessment. *Journal of Psychoeducational Assessment*, *28*(2), 139–152. https://doi.org/10.1177/0734282909344205