

Development of the Short Critical Consciousness Scale (ShoCCS)

Matthew A. Diemer

University of Michigan

Michael B. Frisby

Georgia State University

Andres Pinedo, Emanuele Bardelli, Erin Elliot

University of Michigan

Elise Harris & Sara McAlister

New York University

Adam M. Voight

Cleveland State University

Citation: Diemer, M.A., Frisby, M.B., Pinedo, A., Bardelli, E., Elliot, E., Harris, E.M., McAlister, S. & Voight, A.M. (2022). Development of the Short Critical Consciousness Scale (ShoCCS). *Applied Developmental Science, 26*(3), 409-425. [doi: 10.1080/10888691.2020.1834394](https://doi.org/10.1080/10888691.2020.1834394)

Funding: This project was made possible by:

Institute of Education Sciences Award #R305A170639

William T. Grant Foundation Award #186645

Spencer Foundation Award #201600099

Abstract

Critical consciousness represents the analysis and critique of structural inequalities, the motivation and perceived capacity to effect change, and social action to redress inequity. A wave of recent instruments measuring critical consciousness have been rigorously validated. Yet, whether these measures efficiently assess different levels of critical consciousness or contain redundant, or non-informative, items remains unknown. This research develops a short version (Short CCS, or ShoCCS) of the previously validated Critical Consciousness Scale. Using Item Response Theory methods, the long-form CCS is scrutinized for redundant items to efficiently measure critical reflection, critical motivation, and critical action. The resulting 13-item ShoCCS yields an internally consistent instrument with similar information distributions as the longer measure. Since the ShoCCS places less financial and time burden on investigators and respondents in comparison to the original CCS, and includes a new measure of critical motivation, it may streamline CC measurement in scholarship, practice, and policy.

Development of the Short Critical Consciousness Scale (ShoCCS)

Critical consciousness (CC) represents marginalized people's critical analysis of social injustice as well as motivation and action to challenge social injustices (Watts, Diemer, & Voight, 2011). CC provides people who experience marginalizing social conditions with a more complex and nuanced understanding of the structures that constrain their lives; this newfound understanding facilitates their ability and motivation to negotiate and challenge these constraints in order to self-determine their lives (Freire, 1970). Therefore, CC presents an internal resource, or "psychological armor," that people who are oppressed and/or marginalized can marshal to negotiate and challenge interpersonal, structural, and institutional obstacles (Seider & Graves, 2020).

Canonical conceptions of CC posit three components: (1) critical reflection, the structural awareness of social inequality and the ways in which historical processes perpetuate modern day disparities; (2) critical motivation, individuals' perceived ability and responsibility to enact social change, and (3) critical action, the sociopolitical action taken to rectify inequality (Diemer et al., 2016). Within the CC literature, critical motivation is also referred to as political efficacy and is analogous to sociopolitical control in the empowerment literature (Christens & Dolan, 2011). Higher levels of critical reflection are reflected by making more structural and systemic attributions for societal inequality; lower levels of critical reflection are reflected by making more individualistic attributions for societal inequality (Watts et al., 2011). Higher levels of critical motivation are reflected by higher levels of the perceived capacity and motivation to engage in social activism; lower levels of critical motivation are reflected by lower levels of the perceived capacity and motivation to engage in social activism (Seider & Graves, 2020). Higher levels of critical action are reflected by more frequent and sustained participation in social

activism; lower levels of critical action are reflected by less frequent social activism (McWhirter & McWhirter, 2016).

Higher and lower levels of each CC component can also be illustrated via an example. If presented with information about racial disparities in COVID 19 cases, an adolescent with greater levels of critical reflection would likely point out that people of color disproportionately serve as essential workers because of racism in job opportunities, rather than attribute disparities as caused by individual practices such as smoking. An adolescent who possesses greater levels of critical motivation would experience a sense of urgency to engage in action to alleviate disparities, whereas if they possess low critical motivation, they would feel that action is pointless because nothing will come of it. Lastly, greater critical action could entail organizing with a community group to ensure that businesses have strict safety plans for essential workers.

These components are theorized to be in concert with one another and develop in a reciprocal manner (Seider & Graves, 2020). For example, as youth act to redress perceived inequalities, they learn to see the world in new ways, which leads to growth in critical reflection, which in turn spurs further action. Further, youth's successes and setbacks when engaging in action may foster or diminish (respectively) their critical motivation. Together, these examples highlight how the different facets of CC may inform one another, in a continuous cycle.

CC has been associated with a host of positive outcomes among marginalized youth, primarily in the U.S., including community engagement (e.g., Christens & Dolan, 2011; Pérez-Gualdrón & Helms, 2017), academic achievement (e.g., Cabrera et al., 2014; Seider et al., 2019), healthier sexual decision-making (e.g., Campbell & MacPhail, 2002; Jearey-Graham & Macleod, 2017), and the "5 Cs" of positive youth development (e.g., Clonan-Roy et al., 2016; Delia & Krasny, 2018). Additionally, when measured in adolescence across the transition to adulthood,

CC was associated with attaining higher-status/higher-paying occupations in adulthood, over and above the impact of academic achievement, among marginalized youth in the U.S. (Diemer, 2009; Rapa et al., 2018).

CC scholarship has branched out from its roots in education, to inform medicine (e.g., Dao et al., 2017; Kumagai & Lybson, 2009), public health (e.g., Campbell & MacPhail, 2002; Jearey-Graham & Macleod, 2017), and models of career development (Duffy et al., 2016). Interest in CC has also bloomed internationally, as one measure (the Critical Consciousness Scale, or CCS), has been adapted and/or translated for use in Brazil, Cameroon, France, India, Israel, New Zealand, Romania, South Korea, Sub-Saharan Africa, Switzerland, Ukraine, Zambia, among other nations (M. Diemer, personal communication, February 19, 2020).

A number of recent efforts develop and validate measures of critical consciousness, including the Critical Consciousness Inventory (CCI; Thomas et al., 2014), Measure of Adolescent Critical Consciousness (MACC; McWhirter & McWhirter, 2016), Contemporary Critical Consciousness Measure (CCCM; Shin et al., 2016), Sociopolitical Consciousness (Baker & Brookins, 2014), and the Critical Consciousness Scale (CCS; Diemer, Rapa, Park & Perry, 2017). [Apart from the Sociopolitical Consciousness measure, these were validated with marginalized youth in the U.S.; an overview of and specifics regarding each measure are reviewed in Diemer et al., 2015.] Because CC was previously believed to be too diffuse, abstract, and philosophical to measure via Likert-type items, this spate of measurements is noteworthy and represents an important scholarly advance (Diemer et al., 2015). This paper focuses on one of these measures, the CCS.

Using Item Response Theory Methods to Improve Critical Consciousness Measurement

Recent CC measures, including the CCS, should be applauded for using a variety of rigorous psychometric tools (e.g., exploratory and confirmatory factor analyses carried out with independent samples, estimation of inter-item correlations, using interviews, focus groups, and cognitive interviewing to derive participant perspectives on CC items).

However, existing measurement of CC, including the CCS, could be improved by harnessing the affordances of item response theory (IRT) methods, in at least two important ways: (1) understanding whether items provide information (from an IRT perspective, precision in measurement) about participants at different levels of CC, and (2) removing redundant or non-informative items to yield streamlined CC measures. First, CC measurement could be improved by understanding which items more efficiently provide information about participants with low, medium, or high levels of CC. The factor analytic approach in recently validated CC measures (e.g., CCS, CCCM, MACC) does not indicate which items are more efficient for measuring CC among respondents who, in fact, have different levels of CC. Factor analytic approaches assume that each item contributes an equal amount of “information” about participants (Kline, 2015). That is, EFA and CFA approaches, essentially, assume that each item works equally well and is equally precise for respondents with low, medium, or high levels of CC – and that each item tells us the same amount of information about respondents’ level of CC (DeVellis, 2016; Furr, 2017). Consequently, it is unknown which items and/or measures are better-suited to measure CC among respondents with lower or higher levels of CC – or if all CC measures, potentially, are only informative for respondents with high (or, low) levels of CC.

In contrast, IRT methods estimate how precisely a given item measures the underlying latent construct (i.e., CC) across people with different levels of that latent construct, or to what extent that item provides “information.” Within an IRT framework, items can contribute more or

less information (i.e., measurement precision) about the underlying latent construct, instead of (in an EFA or CFA approach) assuming the amount of information each item contributes is equal (DeVellis, 2016). Specifically, item information curves (IICs) indicate whether an item provides more or less information for respondents across different levels of the underlying latent construct (e.g., critical reflection; Furr, 2017). As Edelen and Reeve (2007) argue: “This feature of IRT [information] is used to evaluate the performance of items and sets of items, and is very useful in constructing short forms or tailored assessments, ensuring that the selected subset of items provide adequate precision across the entire range of interest as well as maximizing precision along critical segments of the construct continuum” (p. 6).

For example, one CCS item may provide little information about respondents with lower levels of the underlying critical reflection latent construct, yet may provide a good deal of information about respondents with higher levels of the underlying critical reflection latent construct. (In a sense, these would represent more “difficult” items in that the respondent would need greater levels of CC in order to endorse that item.) In contrast, a different CCS item (or, set of CCS items) may provide a great deal of information about respondents with lower levels of the underlying critical reflection latent construct, yet provide little information about respondents with higher levels of the underlying critical reflection latent construct. (By extension, these would represent “easier” items in that the respondent would not need greater levels of the underlying CC construct in order to endorse these items).

In sum, the capacity to estimate “information” for each item, and cumulatively, the information for subscales of the CCS, provided by an IRT approach indicate whether a given subscale does well to measure CC only at particular levels of CC or across different levels of CC.

In turn, scrutinizing the CCS will indicate how well that instrument measures CC among respondents with different levels of CC.

The second important way in which CC measurement can be improved with an IRT approach is to identify redundant items so that shorter, yet still comprehensive CC measures can be created. For example, if CCS items #2 and #6 both provide the same levels of information about respondents with higher levels of CC, then the inclusion of both items may introduce redundancy into the CCS measure. IRT approaches provide item information curves (IICs) to inform removal of redundant items. Reducing scale length is important because each item in a survey protocol costs additional money (e.g., participant incentives and costs of survey administration are greater for longer surveys) and time (e.g., each additional item lengthens the survey protocol). Streamlining measures in this way reduces the length of a measure without sacrificing the information that it provides (DeVellis, 2016; Furr, 2017).

Both financial and time costs may compromise the quality of resulting data. Longer measures may inhibit data quality, as participants tire, lose attention, or think less carefully about their responses later in a survey protocol (Little & Rhemtulla, 2013). Longer CC measures may inhibit their uptake and utilization – or their inclusion may necessitate forced exclusion of some other measure(s) of interest from a survey protocol. Researchers must balance competing pressures of time, money, and scientific interests to measure a suite of constructs when designing survey protocols. A shorter CC measure would help researchers more effectively balance these competing demands.

Therefore, the central aim of this research is to develop a short version of the CCS (Short CCS, or ShoCCS). Using an IRT approach, the long-form CCS measure is scrutinized for low-information and redundant items, in order to create a brief CC measure that efficiently measures

critical reflection and critical action. Further, the ShoCCS attempts to incorporate a subscale measuring critical motivation. Items to measure critical motivation were developed for inclusion in the original CCS, yet this subscale was not previously validated with factor analytic or IRT approaches (see Diemer et al., 2017). Subsequent work administered these motivation items to a sample of young people, finding that they formed an internally consistent subscale that purports to measure critical motivation (Rapa et al., 2020).

The current study was divided into two phases, an exploratory (Study 1) and confirmatory (Study 2) phase. The results from Study 1 were used to develop a proposed ShoCCS; Study 2 aimed to replicate this short form of the CCS, with a new source of data. We preregistered Study 2 on the Open Science Framework (<https://osf.io/3qu59>); specifically, we used the results from Study 1 to inform the preregistration plan for Study 2, in the interest of promoting rigor and scientific transparency. Thus, we pre-registered data sources, data inclusion criteria, method for handling missing data, the analytic plan (e.g., type of IRT model—2PL model), and model fit criteria, prior to the Study 2 data analysis. While the general plan was preregistered, IRT models are rare in scale reduction studies and so deviations from the pre-registration were expected. That is, with little prior research utilizing an IRT approach for scale reduction in psychological measurement, it was difficult to anticipate all issues that could arise.

Method

Participants

Data from this study come from three distinct studies examining developmental processes among marginalized youth, each of which administered the CCS. The data utilized in this study will be made available in a repository at the Inter-university Consortium for Political and Social

Research (ICPSR) after an embargo period, following the conclusion of these ongoing data collections.

Study 1 Sample

The first study surveyed 237 youth in a Midwestern city in the U.S (see Diemer, Voight, Marchand and Bañales (2019) for a more comprehensive description of the sample and procedures). The second study surveyed 159 youth from several large metropolitan areas across the U.S. The third study surveyed 165 youth in a large U.S. city on the East Coast. Taken collectively, these data sources offer the advantage of geographic diversity among their participants. For the purposes of the current study, these samples of 159 and 165 youth were combined to yield a sample large enough ($N = 324$) for the IRT analyses, detailed below, in Study 2. We refer to analyses on the 237 participants as Study 1, and analyses on the set of 324 as Study 2.

The 237 youth in Study 1 were public high school students from a Midwestern city, and were mostly seniors (51.9%). The mean age of these students was 16.9 years ($M_{age} = 16.91$, $SD_{age} = 1.25$), and approximately 55.7% identified as female. Nearly half of the students said the highest educational attainment attained by their most educated parent was a high school diploma or less (44.7%), while the rest reported their most educated parent having anything beyond a high school diploma (45.2%). Parent education information was not reported by 10.1% of students. The majority of youth (77.6%) identified as a student of Color (American Indian or Alaskan Native = 0.4%, Asian or Pacific Islander = 2.5%, Black or African American = 58.6%, two or more races = 11.8%, non-white Latino/a = 4.3%). The remaining 22.4% of students identified as White.

Study 2 Sample

The 324 students in Study 2 were predominantly middle and high school students. Two students (0.6%) reported being in the 6th grade, twenty-three (7.1%) in the 7th grade, thirty-three (10.2%) in the 8th grade, fifty-five (17%) in the 9th grade, thirty-four (10.5%) in the 10th grade, fifty-two (16%) in the 11th grade, eighty-seven (26.9%) in the 12th grade, three (0.9%) post-high school, and thirty-five (10.8%) did not offer their grade level. The majority of students identified as female (56.8%), 29.6% identified as male, 0.9% preferred not to respond, 1.9% preferred to describe their gender, and 10.8% offered no response. Most students identified as a student of Color (Asian = 12.3%, African American = 22.5%, Latino/Hispanic = 39.2%, two or more races = 9.6%) The remaining students identified as White (6.3%) or offered no response (10.2%).

Measures

Each Study analyzed items from the original, long-form, CCS (Diemer et al., 2017). Study 1 analyzed 17 items, which comprise the 8-item Critical Reflection: Perceived Inequality and 9-item Critical Action: Sociopolitical Participation subscales. Study 2 analyzed 26 items – these same two subscales – as well as the 9-item Critical Motivation subscale, which was not administered to participants in Study 1. The full set of CCS items are listed in the source publication (Diemer et al., 2017) noted above. A sample item that measures Critical Reflection: Perceived Inequality reads “Certain racial or ethnic groups have fewer chances to get good jobs.” A sample item that measures Critical Motivation reads “It is my responsibility to get involved and make things better for society.” A sample Critical Action: Sociopolitical Participation item reads “[I] Joined in a protest march, political demonstration, or political meeting.” Please also see the supplemental user’s guide, which includes all items that comprise the ShoCCS, for further detail.

These analyses excluded the Critical Reflection: Egalitarianism subscale from the CCS, because Egalitarianism demonstrated a pattern of unexpectedly negative or weak correlations with other CCS subscales (ranging from $-.42$ to $-.10$ in Diemer et al., 2017). Other empirical work also observed weak or negative relations (i.e., ranging from $-.29$ to $-.06$) between a measure of Egalitarianism and other indices of CC (Diemer & Rapa, 2016). [Speculatively, these divergences may be due to the fact that Egalitarianism measures how the world aspirationally “ought to be,” while other CC subscales measure how the world realistically “is.”] Further, the Egalitarianism subscale is only comprised of five items, lessening concerns about the length of this subscale.

Missing data for both studies were addressed via Full Information Maximum Likelihood (FIML), the default in the MPlus software employed for the IRT and CFA analyses. The minimal amount of missing data in each study and the absence of missing data patterns suggestive of problematic and non-ignorable missingness (i.e., MNAR; see Kline, 2015) suggest that the missing data were best classified as missing at random (MAR). FIML makes use of all existing data points, instead of excluding missing data pairwise or listwise, and is well-suited for analyses where missing data are MAR in that it maximizes the information extracted from data that is present (Little & Rhemtulla, 2013). In Study One, 95.4% of participants had complete data for the Critical Reflection subscale and 91.2% for the Critical Action subscale. In Study Two, each subscale had at least 96% complete data.

Analytic Approach

The factor structure of the CCS is well-known, as reported in the original validation paper (Diemer et al., 2017) and replicated in a series of subsequent papers with marginalized youth (e.g., Diemer et al., 2019; Marchand et al., 2019). This strong evidence of

unidimensionality informs the specification of unidimensional IRT models here, where a separate IRT model is fit for each CCS subscale of interest: Critical Reflection: Perceived Inequality, Critical Motivation (in Study 2), and Critical Action.

IRT serves as the overarching analytic approach for Study 1 and Study 2. Study 1 applies IRT methods to each CCS subscale, yielding a short measure that is then replicated via IRT methods with a separate sample in Study 2. More specifically, the Graded Response Model (GRM) is designed for items with ordinal (Likert-type) response options; the CCS contains five- and six-category Likert-type responses (Furr, 2017). Specifically, we use a two parameter logistic (2PL) GRM model with the MLR estimator and logit link function to estimate difficulty (b , or location parameter, in IRT parlance) and discrimination (a , or slope parameter, in IRT parlance) for each CCS item. The “difficulty parameter,” or location, in a GRM model refers to the empirically estimated level of the underlying latent construct (e.g., critical motivation) necessary to cross the threshold into a higher response categories for a Likert-type item, at which a participant would “strongly agree” vs “agree” (for example) with a given item (DeVellis, 2016; Edelen & Reeve, 2007). For example, a more empirically “difficult” CCS item that purports to measure critical motivation reads “People like me should participate in the political activity and decision making of our country” while a less empirically “difficult” CCS item that purports to measure critical motivation reads “It is important for young people to know what is going on in the world.” The “discrimination parameter,” or slope, in a GRM model refers to the degree to which an item can distinguish between respondents with lower vs higher levels at a given range of the underlying latent construct, Theta in IRT parlance. (This is a departure from the more well-known Rasch model, which assumes that each item – or category of response – equally discriminates between people with lower and higher levels of the latent construct; Furr, 2017).

For example, the item reading “[I] contacted a public official by phone, mail, or email to tell him/her how you felt about a social or political issue” was better able to distinguish among youth with higher levels of critical action.

With the aim of reducing an existing scale, the most useful estimates provided by a 2PL model are the item information functions, which are used to create item information curves. These indicate the amount of “information” an item provides about respondents with different levels of the underlying latent construct. Information from an IRT perspective corresponds to precision in measurement; these information curves illuminate at what levels of the underlying latent construct a given item, or subscale, is more precise (Furr, 2017). In these analyses, item information curves would depict how precise Critical Reflection items are for respondents with low vs high levels of critical reflection. For these analyses, these estimates and curves were used to identify redundant items and/or low-information items. Further, item information functions can be aggregated into “test information curves,” which depict information that a given CCS subscale contributes about respondents with varying levels of CC. For example, test information curves here would illuminate how precise the Critical Reflection subscale is for respondents with lower vs higher levels of critical reflection.

Mplus version 8.2 (Muthén & Muthén, 2017) was used for all IRT analyses, including plotting of IICs. R version 3.4.4 was used to graph and compare test information curves. Results were also replicated and confirmed using Stata 16. Test information curves — which graphically illustrate the cumulative information offered by the set of items over a range of Theta—were visually and numerically compared between the full CCS and the ShoCCS. Items were selected for inclusion in the ShoCCS if the test information curves for ShoCCS subscales yielded comparable information about participants as the full CCS. Minimally acceptable information

thresholds were set to 2.33, which corresponds to a reliability estimate of approximately 0.7 (De Ayala, 2013). In addition, we balanced these empirical criteria with substantive judgment, in deriving the short-form of each ShoCCS subscale (e.g., the various forms of critical action measured by the original CCS Critical Action subscale, such as protesting, attending demonstrations, writing public officials, and participating in action groups, were all accounted for by the items that comprise the Critical Action subscale in the ShoCCS). That is, theory was used to decide between items that provided similar levels of information and, in some cases, items that contributed slightly less information were chosen to more broadly measure each component of CC.

We estimate the internal consistency reliability for each subscale using coefficient alpha and we set the cutoff for acceptable reliability to 0.70 (absent a clear consensus on acceptable reliability, the commonly used threshold of .70 was used, see McNeish, 2018). We also report the mean inter-item correlations (IICs) and Omega Total coefficients. These alternative reliability estimates could be better estimators of the measure reliability when the assumptions for coefficient alpha are not met. Specifically, the alpha coefficient is biased against shorter measures (McNeish, 2018) which could lead to underestimation of the reliability of the ShoCCS. Omega Total is preferable to alpha when test items have different factor loadings or discrimination coefficients (Raykov, 1997). However, given the prevalence of alpha in previous literature, we include it for consistency and comparison with other research. Calculation of Omega Total requires factor loadings to be estimated from a confirmatory factor analysis (CFA). Although the factor structure of these measures is well-known (or, suggested via IRT analyses for the Critical Motivation subscale), we will conduct a CFA for the ShoCCS measures in Study 1 and Study 2, and use the fit of the model as a check of concurrent validity.

The aim is to yield at least four items per ShoCCS subscale, to ensure that any latent construct can serve as a standalone latent variable (i.e., for identification purposes) in future research that uses SEM. IICs, coefficient alpha, Omega Total, and CFA fit statistics for the ShoCCS will be contrasted with these same estimates for the full CCS instrument.

Results

Study 1

For a 2PL Graded Response Model (GRM) model, scholars suggest a minimum of 250 participants (Reeve & Fayers, 2005, p. 71). Study 1 is comprised of 237 participants. Therefore, Study 1 may not provide enough participants to produce a stable set of estimates – underscoring the need to replicate the model obtained via Study 1 in a new, and larger, dataset in Study 2. However, the model of interest only analyzes 17 items – generally, IRT analyses are carried out on longer scales, with more items - so Study 1 would require fewer participants than the typical GRM analyses of a longer scale (Furr, 2017).

Visualization of findings across Studies 1 and 2 can be found below in Figures 1-3. Readers who prefer a visual guide to the corresponding narrative of study results are encouraged to look ahead to these figures.

We applied information benchmarks—with information written as $I(\theta)$, where θ refers to “theta”, the underlying latent trait being measured (such as critical reflection)—based on De Ayala (2013). A reliability of at least 0.7 corresponds to $I(\theta) \geq 2.33$, of at least 0.8 corresponds to $I(\theta) \geq 4$, and of at least 0.9 corresponds to $I(\theta) \geq 9$. We report both coefficient alpha and Omega Total as measures of reliability, and use fit statistics based on Hu and Bentler (1999) to establish construct validity ($CFI \geq 0.95$, $TLI = \geq 0.95$, $RMSEA \leq 0.06$).

Graded Response Model: Critical Reflection

Beginning with critical reflection, the GRM indicates that the long-form CCS subscale provides a high degree of information about youth, across a wide range of the underlying critical reflection construct. That is, the CCS is appropriate for youth with low, average, and high levels of critical reflection. Reliability was above .90 (corresponding to $I(\theta) > 9$), for youth within ± 1.5 standard deviations of the average level of critical reflection. Reliability was above .70 for youth at ± 2.5 standard deviations (corresponding to $I(\theta) > 2.33$).

The 8-item Critical Reflection measure was reduced to a 4-item measure, with the 4-item ShoCCS measure providing similar information across levels of critical reflection to the full CCS. The 4-item ShoCCS measure of Critical Reflection exhibited reliability of at least .89 for youth within ± 1.5 standard deviations from average levels of critical reflection ($I(\theta) \geq 8.5$), and had a correlation of 0.97 with its long-form counterpart. Loss of information came at extreme levels of critical reflection; the ShoCCS Critical Reflection subscale does not as precisely measure critical reflection among youth with very low or very high levels of this underlying latent construct. The ShoCCS Critical Reflection subscale is less reliable (0.66) for youth at 2.5 standard deviations below the mean, and for youth 2.5 standard deviations above the mean (0.71). For youth at ± 2 standard deviations from the mean, reliability remains above 0.8.

Figure 1 depicts the test information curves for the original CCS Critical Reflection measure, the ShoCCS measure in Study 1, and the ShoCCS measure in Study 2. The horizontal line reflects the 0.7 reliability threshold. Despite some variability at higher levels of critical reflection, results suggest that the ShoCCS Critical Reflection measure offers reliable information across a wide range of levels of critical reflection.

[INSERT FIGURE 1 ABOUT HERE]

Graded Response Model: Critical Action

Next, the GRM evaluated the informational properties of the full 9-item CCS Critical Action: Sociopolitical Participation subscale. We found that this subscale offers more information for youth with relatively high levels of the underlying critical action latent construct, and less information for youth with average or lower levels of the underlying construct. Information was maximized ($I(\theta) = 11.4$) for youth 1.5 standard deviations above average levels of critical action (this corresponds to a reliability ≥ 0.9). Yet despite the information curve favoring high-critical action youth, the CCS Critical Action scale measures participants with average or above levels of critical action with a reliability of at least 0.87 ($I(\theta) \geq 6.5$).

The full Critical Action scale was reducible to a 5-item measure, with the 5-item measure retaining most of the precision of the longer 9-item measure. The 5-item ShoCCS measure of critical action offered considerably more information for youth with higher levels of critical action, and less information for youth with lower levels on the construct. Information was again maximized (9.7) for participants roughly 1.5 standard deviations above average levels of critical action (this corresponds to a reliability ≥ 0.90). The critical action ShoCCS factor had a correlation of 0.93 with its long-form counterpart. Similar to the information curve for the full CCS Critical Action measure, the information curve for ShoCCS critical action suggests that the items exhibit at least 0.81 reliability for youth with at least average levels of critical action (corresponding to $I(\theta) \geq 4.4$).

The results of these analyses are depicted and tabulated, below. Figures 1 and 2 depict the full CCS and ShoCCS test information curves (which depict how much information the entire measure yields, as opposed to how much information is provided in individual item curves) for Critical Reflection and Critical Action. Results from these analyses informed the factor structure, reliability benchmarks, and general expectations to be replicated in Study 2. That is, we

hypothesized that the same set of items that comprised the shortened scales in Study 1 would also comprise the shortened scales in Study 2, and that a similar degree of reliability would emerge in Study 2.

Study 2

Study 2 is comprised of 324 participants, which exceeds recommendations ($N = 250$) for a 2PL Graded Response Model (GRM; Reeve & Fayers, 2005, p. 71). The aim of Study 2 is to confirm the ShoCCS Critical Reflection and Critical Action measures from Study 1 with a new sample, and to propose a short scale for Critical Motivation. Because Study 2 also analyzes the Critical Motivation subscale (which was not administered to Study 1 participants), this analysis is of 26 items, in total. Descriptive statistics for all items in the ShoCCS measure are provided below in Table 1.

[INSERT TABLE 1 ABOUT HERE]

Graded Response Model: Critical Reflection

Beginning with critical reflection, the 4-item ShoCCS measure provides similar informational properties as Study 1, with a new and geographically diverse sample. The correlation between the ShoCCS Critical Reflection measure and its long-form counterpart was 0.98. With the Study 2 sample, the ShoCCS provides more information than the original CCS and the ShoCCS Study 1 sample at levels of critical reflection between approximately 1.5 standard deviations below average and half a standard deviation above average. (This may be due to some differences across samples – Study 1 participants were older and Study 2 participants were more engaged in youth organizing and youth leadership councils – which we return to in the Discussion.)

For youth one standard deviation above the mean and beyond, the ShoCCS Critical Reflection measure in Study 2 offered less information than in Study 1. However, both studies suggest that the ShoCCS measure offers reliability of at least 0.7 for levels of critical reflection from at least 2 standard deviations below the mean until at least 1.7 standard deviations above the mean. Figure 1 depicts the test information curves for the ShoCCS measure in Study 2, suggesting that the Critical Reflection measure is precise across a wide range of levels of critical reflection.

Graded Response Model: Critical Action

Next, the GRM examined the properties of the ShoCCS Critical Action measure. The ShoCCS Critical Action measure was more precise with the Study 2 sample. Specifically, the measure offered information across a wider range of critical action in Study 2, and showed 0.7 or more reliability throughout this range of critical action. The ShoCCS Critical Action measure correlated with its long-form counterpart at 0.96, and provided the most information for participants with near average levels of critical action compared with the relatively high levels of critical action (+1.5 SD) in Study 1. The measure also exhibited at least 0.7 reliability for youth with critical action levels as low as 1.5 standard deviations below average, versus 0.65 standard deviations below average in Study 1. Moreover, although the ShoCCS Critical Action measure provided less information in Study 2 at higher levels of critical action, the reliability at this end of distribution is still roughly 0.88 or higher until two standard deviations above the mean.

Figure 2, below, provides test information curves for the original CCS measure of Critical Action, the ShoCCS measure of Critical Action in Study 1, and the ShoCCS measure of Critical Action in Study 2. The horizontal line marks the informational threshold for obtaining

reliability of 0.7. In totality, results suggest that the ShoCCS measure of Critical Action provides ample and reliable information for youth with average or above levels of critical action.

[INSERT FIGURE 2 ABOUT HERE]

Graded Response Model: Critical Motivation

Lastly, a GRM examined the properties of the long-form Critical Motivation scale (Rapa et al., in press). The full 9-item measure offers substantially more information for youth with lower levels of critical motivation than average or higher levels. Specifically, information was maximized for participants between 1 and 3 standard deviations below average levels of critical motivation. Average information in this range was approximately 14 ($I(\theta) \approx 14$), corresponding to reliability greater than 0.9. While the information offered by the ShoCCS measure begins to decrease near average levels of critical motivation, it nevertheless remains reliable at 0.7 or more for youth possessing levels of critical motivation up to one standard deviation above the mean.

The 9-item Critical Motivation subscale is reducible to a 4-item subscale that still yields similar amounts of information across similar levels of critical motivation. The short- and long-form measures of critical action correlated at 0.97. Information from the ShoCCS Critical Motivation measure is maximized across the same range of critical motivation (-3SD to -1SD), and also begins to taper for critical motivation levels around the mean. The range of reliability at 0.7 or above is identical for the long and short Critical Motivation measures, although the ShoCCS measure falls below 0.7 reliability just before one standard deviation above average levels, and the full measure just after one standard deviation above average.

Figure 3 depicts test information curves for the full and ShoCCS versions of Critical Motivation. The horizontal line depicts the informational threshold for reliability above 0.7. Writ

large, the ShoCCS Critical Motivation measure exhibits functionality very similar to the full measure.

[INSERT FIGURE 3 ABOUT HERE]

CFA & Omega Total

Following these IRT analyses, confirmatory factor analyses (CFA) were carried out to confirm factor loadings – which were needed to calculate Omega Total, which McNeish (2018) recommends for more precisely estimating reliability. The CFA established that the ShoCCS factor structure held across Studies 1 and 2. The model from Study 1 was a good fit to the data (CFI = .98, TLI = .97, RMSEA = .087, 90% [CI] = [.06, .01], WRMR = .84), although the RMSEA exceeded cutoffs for strong fit. This model fit included covarying the error terms for two pairs of items that likely shared error variance, possibly because of similar wording in the question stems: Q3 and Q4 (“Women have fewer chances to get ahead” and “Poor people have fewer chances to get ahead”; $r = -0.21$) as well as Q5 and Q6 (“I participated in a civil rights group or organization” and “I participated in a political party, club, or organization.”; $r = 0.31$)¹. These covarying error terms were non-significant, but were included to mirror past work with the CCS (see Diemer et al., 2019). Refitting the model without residual correlations yielded a nearly equivalent fit to the data (CFI = .98, TLI = .97, RMSEA = .087, 90% [CI] = [.06, .01], WRMR = .86). RMSEA was above the suggested threshold of 0.06 for a good fit, CFI and TLI were each above their respective good fit thresholds of 0.95, and WRMR was slightly below its target threshold of 1 for good fit (Hu & Bentler, 1999; Kline, 2015). Collectively, this suite of fit indices suggested good model fit; the RMSEA value exceeding the recommended threshold may be due to a smaller sample size in Study 1 (Kline, 2015). Additionally, omega total calculations

¹ These correspond to correlations in the standardized output. The covariances are -0.054 and 0.11, respectively.

suggested that both Critical Reflection and Critical Action exhibited good reliability with this sample ($\omega_{CR} = 0.90$, $\omega_{CA} = 0.78$), reaffirming the reliabilities and strength of the measures indicated by the GRM. These reliability estimates are consistent with coefficient alpha and IICs for each ShoCCS subscale [$\alpha_{CR} = 0.89$, $\alpha_{CA} = 0.80$, $IIC_{CR} = .65$, $IIC_{CA} = .45$].

Study 2 resulted in a model that was an even stronger fit to the data (CFI = .99, TLI = .98, RMSEA = 0.071, 90% [CI] = [.06, .08], WRMR = .87). RMSEA remained marginally above its suggested threshold of 0.06 for a good fit, though the threshold fell within its 90% confidence interval. CFI and TLI were both above their suggested cutoffs for good fit of 0.95, and WRMR was below its recommended cutoff for good fit of 1. This model fit included covarying the error terms for the same two items (listed above) that likely shared error variance based on past research: Q3 and Q4 ($r = .36$); Q5 and Q6 ($r = .32$)². Refitting the model without residual correlations showed a similarly strong fit to the data (CFI = .98, TFI = .98, RMSEA = .077, 90% [CI] = [.06, .09], WRMR = .95). Like Study 1, correlated error terms were included to mirror the structure of the CCS. It should also be noted that Study 2 contains one additional latent construct—critical motivation—that was not present in the CFA for Study 1. Omega total estimates for Critical Reflection, Critical Action, and Critical Motivation suggested strong levels of reliability for each measure with this sample ($\omega_{CR} = .92$, $\omega_{CA} = .80$, $\omega_{CM} = .80$). This is consistent with estimates of coefficient alpha and IICs for each ShoCCS subscale [$\alpha_{CR} = 0.93$, $\alpha_{CA} = 0.82$, $\alpha_{CM} = .80$, $IIC_{CR} = .77$, $IIC_{CA} = .50$, $IIC_{CM} = .48$]. Thus, the CFA and reliability estimates from Study 2 both reaffirm the reliability and strength of the measures identified in the GRM, and demonstrate consistency across geographically diverse samples of youth in the U.S.

Discussion

² These correspond to correlations in the standardized output. The covariances are 0.085 and 0.162, respectively.

Leveraging the affordances of an IRT approach, this research aimed to streamline the CCS. It contributes to the literature in four primary ways. First, Studies 1 and 2 indicate that this short form, the ShoCCS, yields similar (and in some cases, more) information as the long-form CCS. In Study 1, this is indicated by similar theta coverage and reliable information patterns between the ShoCCS and the CCS, while Study 2 shows similar relationships between theta and information with a new data source. Thus, a shortened measure could be reliably and readily adopted, because fewer items pose less financial burden on investigators and less time burden on participants (Little & Rhemtulla, 2013). Studies 1 and 2 indicated, and we subsequently removed, low-information and redundant items. Candidates for item reduction were identified in pre-registration (<https://osf.io/3qu59>), and the decision-making process is supplemented in the Appendix by item information curves, thresholds, and discrimination parameters.

Coefficient alpha, omega total, and IICs indicate that the resulting ShoCCS subscales were quite internally consistent. Additionally, these scales correlated very strongly with the long-form CCS subscales ($0.93 \leq r \leq 0.98$). The resulting ShoCCS is comprised of 13 items that efficiently yield unique information about participants' critical reflection, critical motivation, and critical action (Furr, 2017). The Critical Motivation scale was found to best measure youth with lower levels of critical motivation (-3SD to 0.9SD); the Critical Reflection scale was found to best measure youth near average levels of critical reflection (-2SD to 1.5SD); and the Critical Action scale was found to best measure youth with high levels of critical action (-1SD to 3SD). Notably, there is some variation between the endpoints of theta coverage between studies (particularly at the high end of critical reflection, and the low end of action), as would be expected across samples. All scales were reliable (reliability > 0.7 ; $I(\theta) > 2.33$) across a broad range of theta, including at average levels of its respective domain. More broadly, this intense

psychometric scrutiny further establishes the viability of measuring CC, in general, and provides construct validity evidence in support of the ShoCCS, in particular.

Secondly, this IRT analytic approach provides estimates of how well the ShoCCS measures each domain across a range of the underlying CC construct. In contrast, the factor analytic approach in previous CC instrument validation studies (e.g., Baker & Brookins, 2014; Diemer et al., 2017; McWhirter & McWhirter, 2016; Thomas et al., 2014) assumes that each item contributes an equal amount of information, across the distribution of the underlying CC component being measured (e.g., critical action). This approach, therefore, is not designed to estimate whether the CCS (or, other CC measures) yields more or less information across different levels of CC. In contrast, this IRT analytic approach yields a short form that precisely measures critical reflection, motivation, and action, across different levels of each CC component (Furr, 2017).

Third, the original CCS did not contain a measure of critical motivation, which is argued to be a key component of CC (e.g., Watts et al., 2011). Study 2 provides new and rigorous construct validity evidence for the Critical Motivation subscale, which previously had only been subjected to internal consistency estimates (Rapa et al., in press). Finally, in the interest of transparency and reproducibility, Study 2 was preregistered with the Open Science Framework. Although preregistration can be used to specify hypothesis-generating studies (i.e., exploratory), it is critical that exploratory studies be explicitly differentiated from hypothesis-confirming studies (i.e., confirmatory). Preregistering Study 2, then, ensures that data were not recoded, reanalyzed, or re-specified in order to reach a certain criterion or goal and that our results indeed support confirmatory hypotheses rather than exploratory ones. Additionally, though we had expectations that we may need to deviate from pre-registration plans due to the exploratory

nature and novelty of this study, no such deviations were necessary. Thus while Preregistration of IRT analyses is less common, this paper suggests that it is viable and ought to be more frequent in the literature.

Because CC has been associated with a number of positive developmental outcomes, particularly among marginalized youth (as summarized in Diemer et al., 2016), the ShoCCS provides a streamlined and precise measurement tool to advance existing knowledge, policy and practice. For example, recent inquiry (Seider & Graves, 2020) indicates that different schooling models, such as supporting youths' engagement with the political system and mobilizing to challenge inequalities, are associated with students' growth in specific domains of CC – in this case, critical action. The ShoCCS would provide an efficient measurement strategy to assess initial levels of CC, as well as change over time, that would reduce participant burden and financial cost to investigators over multiple waves of data collection. This may attenuate participant attrition, item non-response, and missing data, while potentially improving the resulting data quality (Little & Rhemtulla, 2013).

Limitations & Future Directions

As noted above, the current study did not include the Critical Reflection: Egalitarianism subscale from the original CCS. Because this subscale has not previously correlated in expected directions with other CC subscales, or demonstrated convergent validity, and because the Egalitarianism subscale is already quite brief – five items – it was not included in the ShoCCS. This brief subscale could be used in the future, but we do not know whether it provides different amounts of information for youth with lower or higher levels of Egalitarianism.

It stands to reason that prior construct validity evidence in support of the CCS would also support the ShoCCS. Yet, future research should provide further construct validity evidence for

the ShoCCS, particularly the Critical Motivation subscale, because that subscale was only examined in Study 2. For example, convergent validity evidence could be obtained by correlating ShoCCS subscales with related subscales from other CC indices, such as the CCCM (Shin et al., 2016), CCI (Thomas et al., 2014) or the MACC (McWhirter & McWhirter, 2016). Similarly, divergent validity evidence could be obtained by correlating the ShoCCS with measures such as the Social Dominance Orientation (Ho et al., 2015) or the Color Blind Racial Attitudes Scale (Neville et al., 2000). Unfortunately, this convergent and divergent validity evidence could not be estimated in this study, because these other measures were not administered to either sample. Finally, test-retest reliability estimates were not obtained in the present study, and would also provide another form of psychometric evidence with which to evaluate the ShoCCS.

Future research should further evaluate the ShoCCS for measurement invariance and differential item functioning (DIF). Within-sample strata limitations prevented multiple group analysis for testing measurement invariance. For example, if the individual samples were split into groups across social identity categories – and even further split to probe intersecting social identity categories – the samples of 237 and 324 would be quite underpowered. Similarly, it may be that the ShoCCS provides more information for Youth of Color than for White youth; sample size precluded examining a distinct GRM for each of these social identities (again, or for intersecting social identity categories).

Previous research (Diemer et al., 2019; Marchand et al., 2020) did not exhibit DIF (across distinct social identity categories of race/ethnicity, gender, or social class) for the Critical Reflection: Perceived Inequality and Critical Action: Sociopolitical Action CCS subscales, respectively, with samples of marginalized youth. Because these two ShoCCS subscales are

streamlined versions of these same two longer subscales, it stands to reason that ShoCCS items would not exhibit DIF across these social identity categories. (The Critical Motivation subscale, however, has not been subjected to the same prior scrutiny and so is more deserving of attention.)

Yet, because is an important open question, we used MIMIC (Multiple Indicator Multiple Cause) models to evaluate DIF on the basis of race/ethnicity and gender, for each ShoCCS subscale. MIMIC models, which are CFA models with exogenous covariates (here, race/ethnicity dichotomized into White youth and Youth of Color and gender dichotomized into male and female) specified to test for latent mean differences and/or DIF (Kline, 2015). In short, we found no evidence in support of DIF for any of the three ShoCCS subscales, along these two identity categories. However, these MIMIC models and previous scrutiny (i.e., Diemer et al., 2019; Marchand et al., 2020) of DIF across race/ethnicity and gender were not able to explore heterogeneity within the more monolithic grouping of youth of color (e.g., examining whether items function differently for Black vs Latinx youth, because sample sizes were too small) or among gender non-conforming youth (because this identity was not measured). Future research should carefully investigate this question, to ensure that the ShoCCS functions equally well across various social identities, as well as their intersections. It will be particularly important to further probe and clarify potential differences at the intersections of these social identity categories, as CC scholarship moves to incorporate intersectionality theory (Godfrey & Burson, 2018).

Future research could create “more difficult” items for Critical Motivation—items which require higher levels of critical motivation in order for participants to endorse them, and thereby differentiate between higher levels of critical motivation. These items would complement the

ShoCCS measure of Critical Motivation, which offered the most information for participants with lower levels of critical motivation. Similarly, future research could create “easier” items for Critical Action—items which require lower levels of critical action in order for participants to endorse them, and thereby differentiate between lower levels of critical action. These items would complement the ShoCCS measure of Critical Action, which offered the most information for participants with higher levels of critical action. These “harder” and “easier” items could be incorporated into a new version of the full CCS, or perhaps a revised ShoCCS.

Longer scales generally capture extreme ranges better than shorter scales, simply because they include more items. The ShoCCS provides less theta coverage at more extreme ranges of critical reflection than the CCS; the ShoCCS provides less theta coverage at extremely high ranges of action yet provides more theta coverage at lower ranges of action than the CCS. The affordances of a brief scale vs. the affordances of a long scale (e.g., more theta coverage) can be considered, and the “right tool for the job” selected. Scale users are encouraged to calibrate their needs to the affordances provided by the long vs short form of this measure.

Conclusion

This paper indicates that CC may be accurately measured, using the ShoCCS, precisely providing levels of information about participants (and in some cases, more information) than the long-form CCS. These measurement properties also held true across multiple samples: urban-residing high school students and a second sample of historically marginalized young people engaged in youth organizing and youth leadership councils, providing evidence for construct validity. The ShoCCS and original CCS subscales were very highly correlated. Further, the ShoCCS efficiently measures each domain of CC - critical reflection, critical motivation, and critical action - across a wide range of CC levels. The Critical Motivation subscale, newly

validated within the ShoCCS, demonstrated strong measurement properties. Given these affordances and insights, the ShoCCS may further improve the measurement of CC in scholarship and practice.

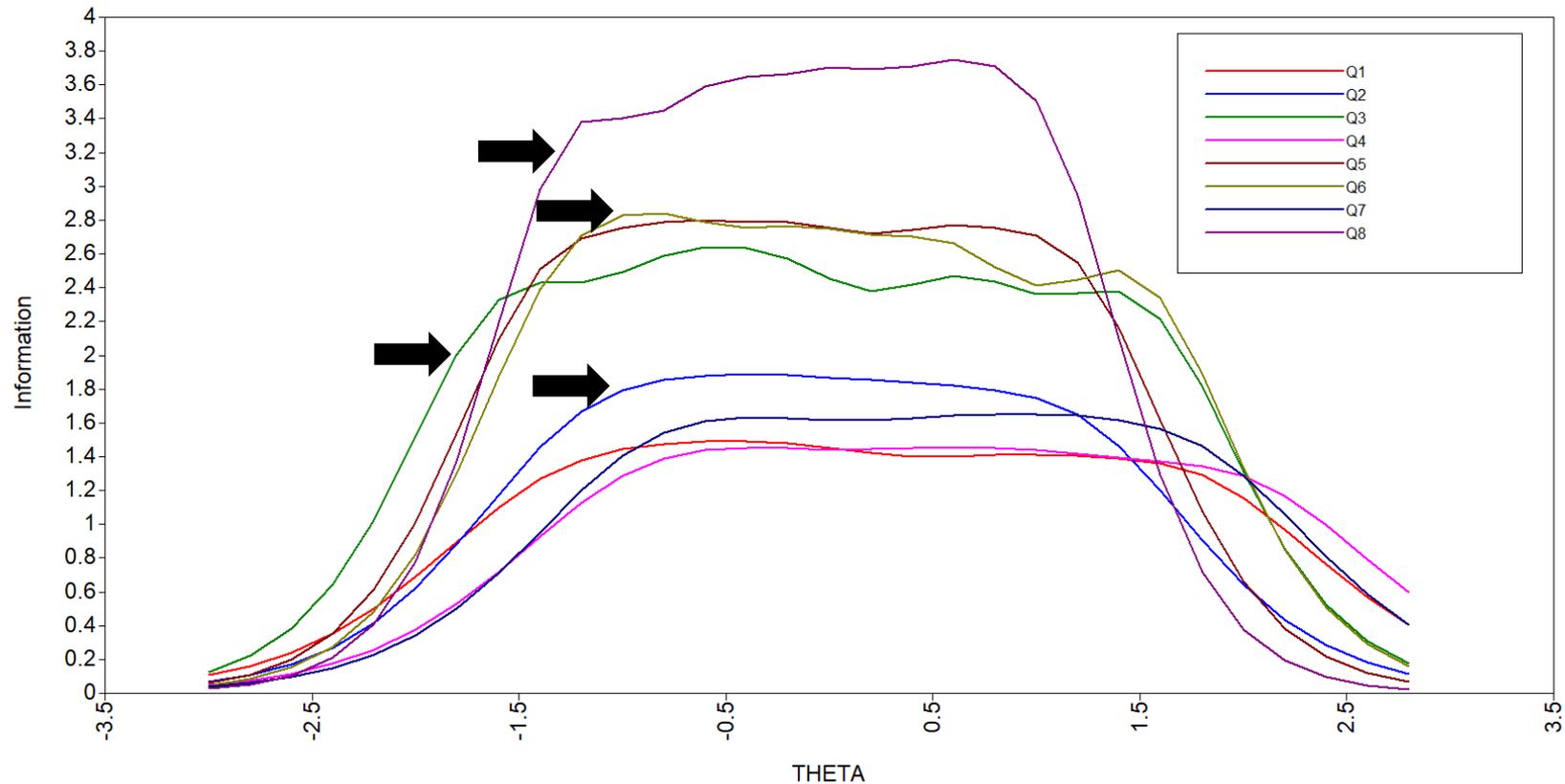
References

- Baker, A. M., & Brookins, C. C. (2014). Toward the development of a measure of sociopolitical consciousness: Listening to the voices of Salvadoran youth. *Journal of Community Psychology, 42*(8), 1015-1032.
- Cabrera, N. L., Milem, J. F., Jaquette, O., & Marx, R. W. (2014). Missing the (student achievement) forest for all the (political) trees: Empiricism and the Mexican American studies controversy in Tucson. *American Educational Research Journal, 51*(6), 1084-1118.
- Campbell, C., & MacPhail, C. (2002). Peer education, gender and the development of critical consciousness: Participatory HIV prevention by South African youth. *Social Science & Medicine, 55*(2), 331–345. [https://doi.org/10.1016/S0277-9536\(01\)00289-1](https://doi.org/10.1016/S0277-9536(01)00289-1)
- Christens, B. D., & Dolan, T. (2011). Interweaving youth development, community development, and social change through youth organizing. *Youth & Society, 43*(2), 528–548. <https://doi.org/10.1177/0044118X10383647>
- Clonan-Roy, K., Jacobs, C. E., & Nakkula, M. J. (2016). Towards a model of positive youth development specific to girls of color: Perspectives on development, resilience, and empowerment. *Gender Issues, 33*(2), 96-121.
- Dao, D. K., Goss, A. L., Hoekzema, A. S., Kelly, L. A., Logan, A. A., Mehta, S. D., ... & DeLisser, H. M. (2017). Integrating theory, content, and method to foster critical consciousness in medical students: a comprehensive model for cultural competence training. *Academic Medicine, 92*(3), 335-344.
- De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.
- Delia, J., & Krasny, M. E. (2018). Cultivating positive youth development, critical consciousness, and authentic care in urban environmental education. *Frontiers in Psychology, 8*, 2340.
- DeVellis, R. F. (2016). *Scale development: Theory and applications*. Sage publications.
- Diemer, M. A. (2009). Pathways to occupational attainment among poor youth of color: The role of sociopolitical development. *The Counseling Psychologist, 37*(1), 6–35. <https://doi.org/10.1177/0011000007309858>
- Diemer, M. A., McWhirter, E. H., Ozer, E. J., & Rapa, L. J. (2015). Advances in the conceptualization and measurement of critical consciousness. *The Urban Review, 47*(5), 809–823. <https://doi.org/10.1007/s11256-015-0336-7>

- Diemer, M. A., Rapa, L. J., Voight, A. M., & McWhirter, E. H. (2016). Critical consciousness: A developmental approach to addressing marginalization and oppression. *Child Development Perspectives*, 10(4), 216–221. <https://doi.org/10.1111/cdep.12193>
- Diemer, M. A., & Rapa, L. J. (2016). Unraveling the complexity of critical consciousness, political efficacy, and political action among marginalized adolescents. *Child Development*, 87(1), 221–238. <https://doi.org/10.1111/cdev.12446>
- Diemer, M. A., Rapa, L. J., Park, C. J., & Perry, J. C. (2017). Development and validation of the Critical Consciousness Scale. *Youth & Society*, 49(4), 461–483. <https://doi.org/10.1177/0044118X14538289>
- Diemer, M. A., Voight, A. M., Marchand, A. D., & Bañales, J. (2019). Political identification, political ideology, and critical social analysis of inequality among marginalized youth. *Developmental Psychology*, 55(3), 538–549. <https://doi.org/10.1037/dev0000559>
- Duffy, R. D., Blustein, D. L., Diemer, M. A., & Autin, K. L. (2016). The Psychology of Working Theory. *Journal of Counseling Psychology*, 63(2), 127-148. <https://doi.org/10.1037/cou0000140>
- Edelen, M., & Reeve, B. (2007). Applying Item Response Theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, 16, 5-18.
- Freire, P. (1970). *Pedagogy of the oppressed* (MB Ramos, Trans.). *New York: Continuum, 2007.*
- Furr, R. M. (2017). *Psychometrics: an introduction*. Sage Publications.
- Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J, Pratto, F., Henkel, K. E., Foels, R., & Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO7 scale. *Journal of Personality and Social Psychology*, 109(6), 1003-1028.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of two- and three-parameter logistic item characteristic curves: A monte carlo study. *Applied Psychological Measurement*, 6(3), 249–260. <https://doi-org.proxy.lib.umich.edu/10.1177/014662168200600301>

- Jearey-Graham, N., & Macleod, C. I. (2017). Gender, dialogue and discursive psychology: a pilot sexuality intervention with South African High-School learners. *Sex Education, 17*(5), 555-570.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.
- Kumagai, A. K., & Lypton, M. L. (2009). Beyond cultural competence: Critical consciousness, social justice, and multicultural education. *Academic Medicine, 84*(6), 782–787. <https://doi.org/10.1097/ACM.0b013e3181a42398>
- Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives, 7*(4), 199–204. <https://doi-org.proxy.lib.umich.edu/10.1111/cdep.12043>
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412–433. <https://doi-org.proxy.lib.umich.edu/10.1037/met0000144>
- McWhirter, E. H., & McWhirter, B. T. (2016). Critical consciousness and vocational development among Latina/o high school youth: Initial development and testing of a measure. *Journal of Career Assessment, 24*(3), 543-558.
- Muthén, L.K. and Muthén, B.O. (1998-2017). *Mplus User's Guide. Eighth Edition*. Los Angeles, CA: Muthén & Muthén.
- Neville, H.A., Lilly, R.L., Duran, G., Lee, R.M. & Browne, L. (2000). Construction and initial validation of the Color-Blind Racial Attitudes Scale (CoBRAS). *Journal of Counseling Psychology, 47*(1), 59–70. <https://doi:10.1037//0022-0167.47.1.59>.
- Pérez-Gualdrón, L., & Helms, J. E. (2017). A longitudinal model of school climate, social justice orientation, and academic outcomes among Latina/o students. *Teachers College Record, 119*(10), 1–37.
- Pratto, F., James S., Stallworth, L.M. & Malle. B.F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology 67*(4), 741-763. <http://dx.doi.org/10.1037/0022-3514.67.4.741>
- Rapa, L. J., Diemer, M. A., & Roseth, C. J. (2020). Can a values-affirmation intervention bolster academic achievement and raise critical consciousness? Results from a small-scale field experiment. *Social Psychology of Education, 23*, 537-557.
- Rapa, L. J., Diemer, M. A., & Bañales, J. (2018). Critical action as a pathway to social mobility among marginalized youth. *Developmental Psychology, 54*(1), 127–137. <https://doi.org/10.1037/dev0000414>

- Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21*(2), 173–184.
- Seider, S. & Graves, D. (2020). *Schooling for critical consciousness: Engaging Black and Latinx youth in analyzing, navigating, and challenging racial injustice*. Cambridge, MA: Harvard University Press.
- Seider, S., Clark, S., & Graves, D. (2019). The development of critical consciousness and its relation to academic achievement in adolescents of color. *Child Development*.
<https://doi.org/10.1111/cdev.13262>
- Shin, R. Q., Ezeofor, I., Smith, L. C., Welch, J. C., & Goodrich, K. M. (2016). The development and validation of the Contemporary Critical Consciousness Measure. *Journal of Counseling Psychology, 63*(2), 210.
- Thomas, A. J., Barrie, R., Brunner, J., Clawson, A., Hewitt, A., Jeremie-Brink, G., & Rowe-Johnson, M. (2014). Assessing critical consciousness in youth and young adults. *Journal of Research on Adolescence, 24*(3), 485-496.
- Watts, R. J., Diemer, M. A., & Voight, A. M. (2011). Critical consciousness: Current status and future directions. *New Directions for Child and Adolescent Development, 2011*(134), 43-57.

APPENDIX**Critical Reflection (ShoCCS in Bold)**

Q1: Certain racial or ethnic groups have fewer chances to get a good high school education.

Q2: Poor children have fewer chances to get a good high school education.

Q3: Certain racial or ethnic groups have fewer chances to get good jobs.

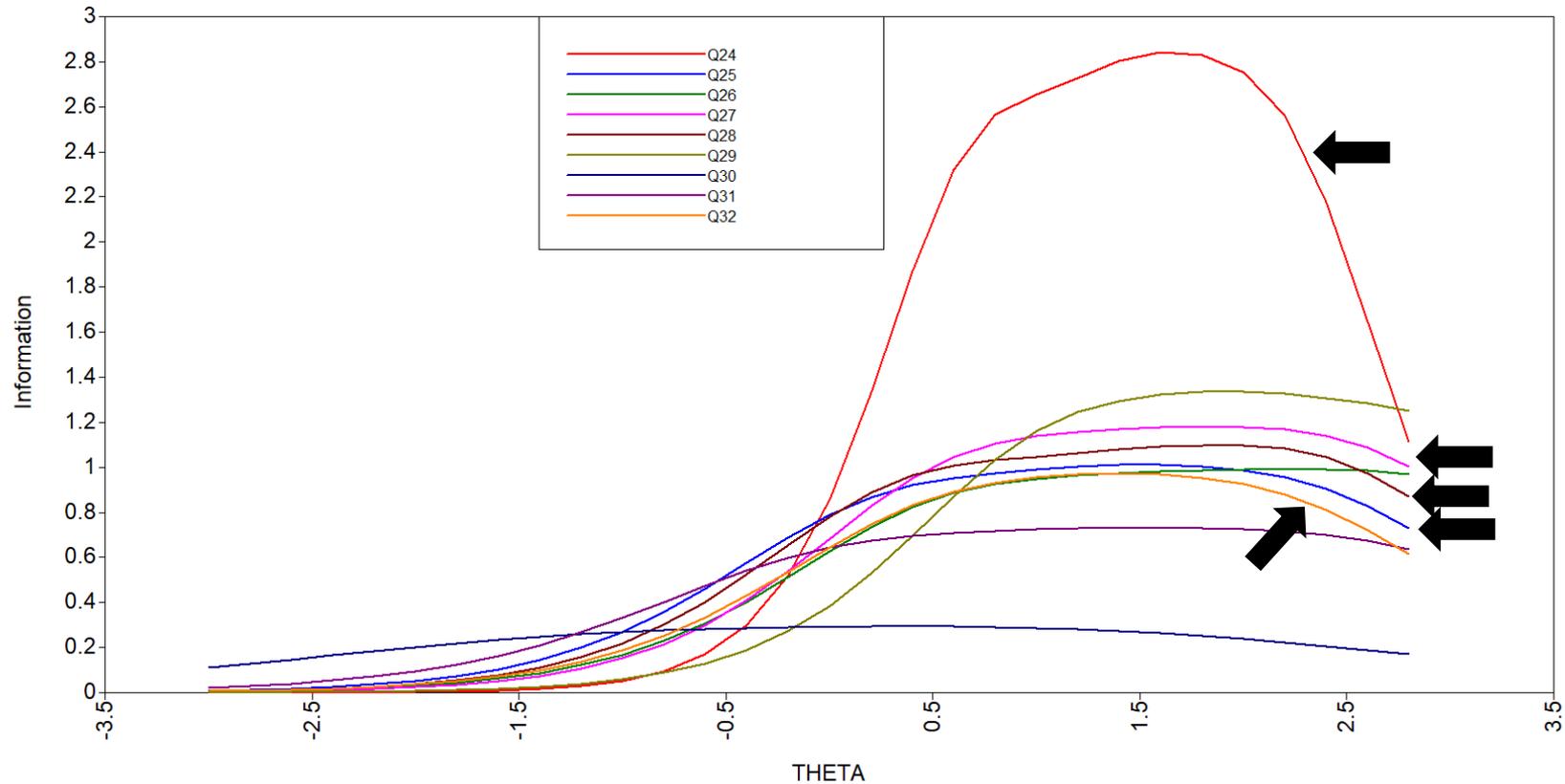
Q4: Women have fewer chances to get good jobs.

Q5: Poor people have fewer chances to good jobs.

Q6: Certain racial or ethnic groups have fewer chances to get ahead.

Q7: Women have fewer chances to get ahead.

Q8: Poor people have fewer chances to get ahead.

Critical Action (ShoCCS in Bold)

Q24: Participated in a civil rights group or organization

Q25: Participated in a political party, club, or organization.

Q26: Wrote a letter to a school or community newspaper or publication about a social or political issue.

Q27: Contacted a public official by phone, mail, or email to tell him/her how you felt about a social or political issue.

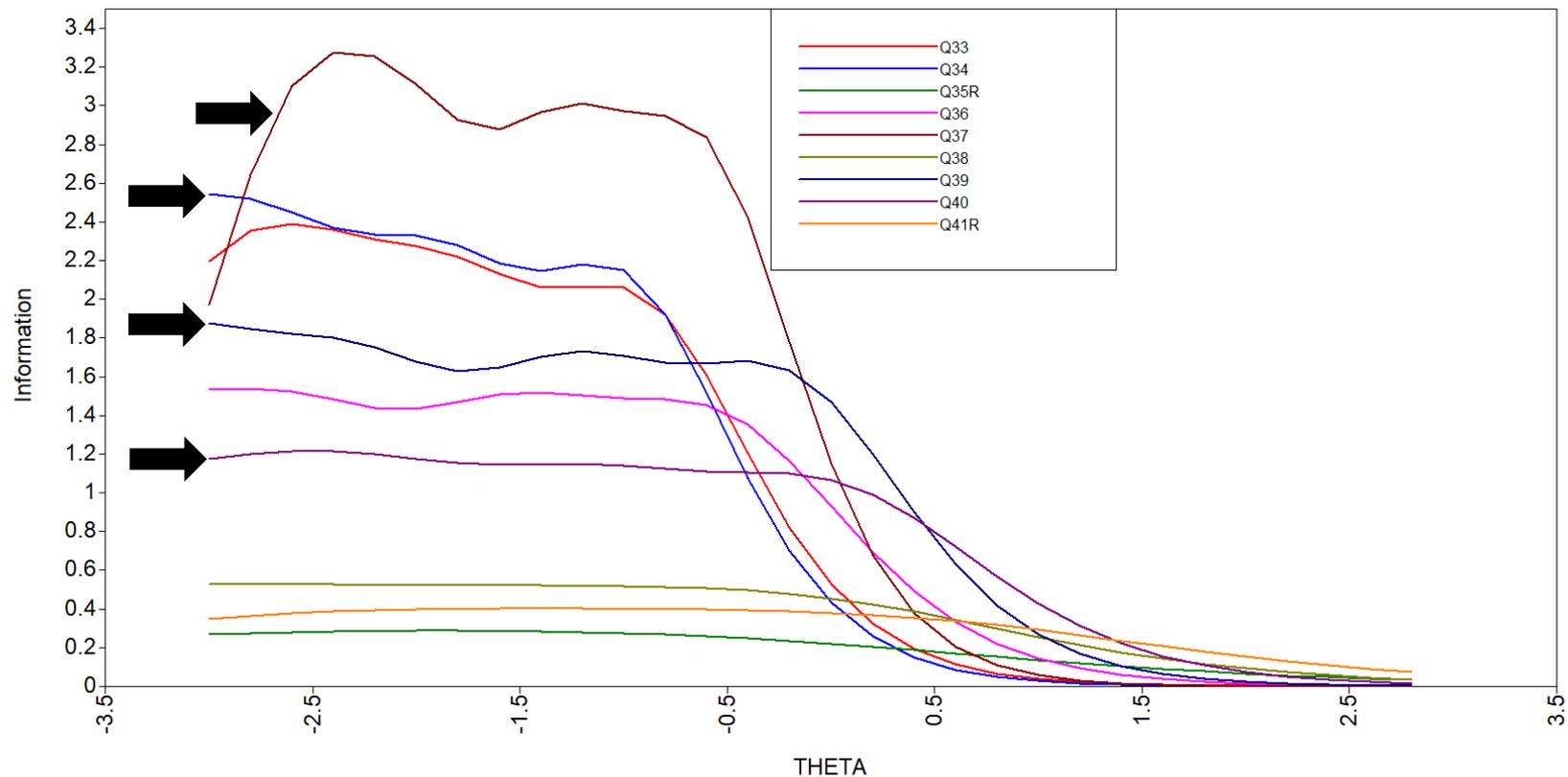
Q28: Joined a protest march, political demonstration, or political meeting

Q29: Worked on a political campaign

Q30: Participated in a discussion about a social or political issue

Q31: Signed an email or written petition about a social or political issue

Q32: Participated in a human rights, gay rights, or women's rights organization or group

Critical Motivation (ShoCCS in Bold)

Q33. Young people have an important role to play in making the world a better place

Q34. It is important for young people to know what's going on in the world.

Q35. Political issues are not relevant to people who are not old enough to vote. (r)

Q36. It is important to be an active and informed citizen.

Q37. It is important to correct social and economic inequality.

Q38. It is important to confront someone who says something that you think is racist or prejudiced.

Q39. It is my responsibility to get involved and make things better for society.

Q40. People like me should participate in the political activity and decision making of our country.

Q41. It does not matter whether I participate in local organizations or political activity because so many other people are involved. (r)

CCS Discrimination and Thresholds (ShoCCS in Bold)

	Discrimination (SE)	Threshold 0-1 (SE)	Threshold 1-2 (SE)	Threshold 2-3 (SE)	Threshold 3-4 (SE)	Threshold 4-5 (SE)
Critical Reflection						
Q1	2.19	-2.80 (0.32)	-1.33 (0.24)	-0.13 (0.22)	1.92 (0.26)	3.84 (0.39)
Q2	2.45	-2.87 (0.33)	-1.53 (0.26)	-0.33 (0.24)	1.16 (0.25)	2.92 (0.33)
Q3	2.96	-4.66 (0.52)	-2.29 (0.33)	-0.66 (0.27)	1.87 (0.32)	4.42 (0.45)
Q4	2.17	-1.99 (0.27)	-0.79 (0.23)	0.90 (0.23)	2.36 (0.29)	4.33 (0.48)
Q5	3.06	-4.15 (0.49)	-2.23 (0.33)	-0.38 (0.28)	1.71 (0.32)	3.62 (0.43)
Q6	3.07	-3.89 (0.51)	-2.36 (0.36)	-0.34 (0.27)	1.76 (0.34)	4.55 (0.61)
Q7	2.31	-2.01 (0.29)	-0.68 (0.24)	1.12 (0.24)	2.53 (0.34)	4.06 (0.50)
Q8	3.53	-4.42 (0.55)	-1.97 (0.35)	0.00 (0.31)	1.93 (0.37)	3.63 (0.48)
Critical Action						
Q24	3.01 (0.54)	2.14 (0.41)	4.08 (0.55)	5.22 (0.75)	6.64 (0.88)	
Q25	1.79 (0.29)	0.49 (0.21)	2.18 (0.30)	2.97 (0.37)	4.29 (0.51)	
Q26	1.79 (0.30)	1.04 (0.23)	2.79 (0.31)	4.26 (0.47)	5.70 (0.64)	
Q27	1.94 (0.31)	1.18 (0.24)	2.70 (0.33)	3.96 (0.42)	5.16 (0.59)	
Q28	1.88 (0.29)	0.76 (0.22)	2.64 (0.31)	3.98 (0.39)	4.72 (0.51)	
Q29	2.06 (0.35)	2.19 (0.33)	3.58 (0.40)	4.59 (0.52)	6.23 (0.85)	
Q30	0.96 (0.16)	-1.08 (0.17)	0.19 (0.16)	0.99 (0.18)	1.51 (0.19)	
Q31	1.52 (0.21)	0.06 (0.19)	1.59 (0.23)	2.84 (0.30)	4.10 (0.46)	
Q32	1.75 (0.26)	0.91 (0.22)	2.14 (0.26)	2.71 (0.30)	3.90 (0.40)	
Critical Motivation						
Q33	2.75 (0.51)	-8.30 (1.41)	-7.71 (1.67)	-6.77 (0.94)	-5.02 (0.69)	-2.52 (0.42)
Q34	2.84 (0.52)	-9.78 (2.62)	-8.76 (1.65)	-7.54 (1.14)	-5.42 (0.72)	-2.82 (0.45)
Q35	0.94 (0.15)	-3.06 (0.31)	-2.58 (0.26)	-2.09 (0.22)	-1.35 (0.17)	-0.54 (0.14)
Q36	2.31 (0.30)	-8.18 (0.87)	-6.10 (0.65)	-3.53 (0.37)	-1.25 (0.24)	-
Q37	3.24 (0.50)	-8.59 (1.28)	-7.86 (1.13)	-6.68 (0.85)	-4.18 (0.56)	-1.96 (0.36)
Q38	1.30 (0.20)	-5.89 (0.66)	-4.51 (0.40)	-3.23 (0.28)	-1.66 (0.18)	-0.19 (0.15)
Q39	2.48 (0.34)	-8.66 (0.95)	-7.66 (0.99)	-5.73 (0.54)	-3.06 (0.34)	-0.55 (0.22)
Q40	1.98 (0.21)	-7.72 (1.20)	-5.63 (0.59)	-4.33 (0.38)	-2.23 (0.24)	0.03 (0.18)
Q41	1.12 (0.15)	-3.19 (0.30)	-2.65 (0.25)	-1.71 (0.19)	-0.74 (0.15)	0.44 (0.14)

Critical Reflection

Similar theta coverage was offered by all critical reflection items. Although there were obvious differences in the amounts of information provided by each item, the information provided by most items was nevertheless acceptable. Question 8, which regarded the opportunities available to the poor, offered the most information over a wide range of theta, and was kept on this basis. Questions 5 and 6 provided nearly identical levels of information over similar ranges of theta (to one another), but question 5 was qualitatively similar to question 8, and thus we opted to omit it in favor of question 6 on this basis. Question 3 offered the next highest level of information, and was qualitatively distinct from question 6 in its focus on jobs rather than opportunities. Of the items remaining, a key qualitative distinction that had not been yet included was an emphasis on women. Between questions 4 and 7 pertaining to women, question 4 was among the lowest information items, and thus we chose question 7 on the basis of it being the most informative question for this qualitative dimension. Notably, there are additional qualitative dimensions that could have been selected on, but in the interest of minimizing the number of items in the scale, we chose to omit them.

Critical Action

The majority of critical action items yielded similar levels of information across similar ranges of theta with the exception of question 24, which asks about participants' involvement in a civil rights group or organization. Question 24 was included because it carries the most information across a similar range of theta. Despite similar relationships between theta and information across other items, these items vary slightly with respect to the amount of information provide and the range of theta covered. Moreover, there are qualitative differences between these items. Question 28 was included, for example, based on its information on involvement in protests and marches. It also covers a moderately wider range of theta than some other items. Similarly, question 27 regards more traditional means of action—contacting public officials—in addition to providing slightly elevated levels of information. By contrast, question 29 was omitted (despite offering the second highest level of information) on the basis of its similarity with question 25, and that it pertained predominantly to individuals with the highest levels of critical action.

Critical Motivation

Among the critical motivation items, question 37 pertaining to correcting social and economic inequality provided the most information at levels of theta comparable to other items in the critical motivation scale. Questions 33 and 34 offered the next highest levels of information, and are nearly identical in their relationships between information and theta. Question 33, however, emphasized making the world a better place, which we felt qualitatively overlaps question 37. Thus, question 34 emphasizing understanding the world felt more qualitatively distinct and offered the same information. Questions 39 and 40 both underscored personal responsibility for involvement, and offered marginally wider ranges of theta coverage than other items. Thus in addition to their psychometric properties, we felt they nicely complemented the other questions included in the critical motivation scale.

Across all scales, we omitted items on the basis that they provided minimal information of the remaining eligible items, and were not qualitatively distinct enough to sacrifice scale brevity. We acknowledge that some of the omitted items have markedly different thresholds than some of the

items retained (which offers another dimension of variability across items), but we did not feel this additional threshold variation warranted their inclusion given the limited information they provided.