**The Support for Economic Inequality Scale: Development and Adjudication**

Dylan Wiwad1, Brett Mercier2, Michael D. Maraun1, Angela R. Robinson2, Paul K. Piff2, Lara B. Aknin1, & Azim F. Shariff2

1Simon Fraser University; 2University of California, Irvine

WORD COUNT: 8,495

Conflict of Interest: The authors certify that they have NO affiliations or involvement in any organization or entity with any financial, or non-financial, interest in the subject matter or materials discussed in this manuscript.

CORRESPONDING AUTHOR:

Dylan Wiwad

Department of Psychology

Simon Fraser University

8888 University Drive

Burnaby, BC

Canada

V5A 1S6

Email: [dwiwad@sfu.ca](mailto:dwiwad@sfu.ca)

Phone: 778-782-9838

**Abstract**

Economic inequality is a defining issue of the 21st century. Past research has documented myriad pernicious psychological effects of high inequality, and prompted interest into how people perceive, evaluate, and react to economic inequality. Here we propose, refine, and validate the Support for Economic Inequality Scale (SEIS) – a novel measure of attitudes towards economic inequality. In Study 1, we distill eighteen items down to five, providing evidence for unidimensionality and reliability. In Study 2, we replicate the scale’s unidimensionality and reliability and demonstrate its validity. In Study 3, we evaluate a United States version of the SEIS. Finally, in Study 4, we show predictive validity by relating the SEIS to behavioral support for an inequality reducing social policy. The SEIS is a valid and reliable instrument for assessing perceptions of and reactions to economic inequality, and provides a useful tool for researchers investigating the psychological underpinnings of economic inequality.

Keywords: Economic inequality, Item Response Theory, Measurement, Scale development

**Support for Economic Inequality (SEIS): Scale Development and Adjudication**

In 2013, President Barack Obama labeled economic inequality one of the most pressing issues of our time (Obama, 2013)—an assertion mounting empirical evidence corroborates (e.g. Pickett & Wilkinson, 2017; Piketty & Saez, 2014). Economic inequality has substantial psychological costs (e.g., Pickett & Wilkinson, 2006; Wilkinson, 2006) and social consequences (e.g., Murphy, 2001), such as decreased emotional well-being, reduced physical health, and decreased social trust (see Buttrick & Oishi, 2017 for a comprehensive overview of the negative consequences of inequality). And yet despite growing evidence that inequality is detrimental to people and societies, in many countries little action is being taken towards addressing these inequalities (Pickett & Wilkinson, 2017). Why? One possibility may lie in people’s attitudes toward economic inequality––the extent to which it is deemed to exist, be problematic, and warrant amelioration. In the current paper, we propose a novel measure of people’s attitudes to inequality by developing and evaluating a short, reliable, and informative measure of support for economic inequality.

Economic inequality has received growing levels of attention in the media and social scientific research. For example, within the last two decades alone, there has been an annual increase in the publication of articles mentioning income inequality or economic inequality, from 67 articles published in 1993 compared with 937 articles published in 2016 – a 1,298% increase (Figure 1; Web of Science). Interest has steadily risen as researchers aim not just to quantify levels of inequality that exist within a society, but understand how inequality is maintained through perceptions of, and attitudes toward, economic inequality (e.g., Engelhardt & Wagener, 2014; Kraus, Piff, & Keltner, 2009; Piff, Kraus, & Keltner, 2017; Norton & Ariely, 2011; Shariff, Wiwad, & Aknin, 2016). In addition, researchers have started exploring the daily experience of inequality, bridging the gap between abstract economic indicators and psychological experience and perception (e.g., Buttrick & Oishi, 2017; Kraus, Park, & Tan, 2017).

Surprisingly, researchers have not yet validated a measure assessing support for economic inequality, though numerous indices capture related constructs. The majority of past research uses existing questions in large cross-national datasets, such as the World Values Survey (e.g., “Incomes should be made more equal” versus “We need larger income differences as incentives for individual effort”), European Values Survey (“Incomes should be made more equal” versus “There should be greater incentives for individual effort;” Kaltenthaler, Ceccoli, & Gelleny, 2008), or International Social Survey Programme (“Differences in income in <respondent’s country> are too large;” Kiatpongsan & Norton, 2014; Shariff, Wiwad, & Aknin, 2016). While face valid and convenient, these items lack empirical psychometric evaluation. Without adjudication, we cannot be sure that items on a scale are assessing the construct they intend to measure, and how well they are doing so. Indeed, recent work demonstrates that proper scale development and construct validation procedures are often ignored in social psychology (Flake, Pek, & Hehman, 2017), despite the fact that the accuracy of psychological measurement, and by extension the validity of psychological results, is entirely dependent on proper scale development and adjudication. Thus, the present paper aims to fill this void by offering a short, reliable, and psychometrically sound scale of support for economic inequality.

**Defining “Support for Economic Inequality”**

A clear definition of the construct in question is crucial to the construction of any measurement tool. We designed the present scale to measure the degree to which one supports or opposes the current level of economic inequality *as they perceive it*. We specify “as they perceive it” because past research has demonstrated that people vary in their perceptions of how much inequality exists, and should exist –– for example, perceptions of the ideal ratio of pay between the average worker and CEO can range from 2.0 to 20.0 (Kiatpongsan & Norton, 2014). Additionally, the scale does not present any factual information regarding the actual level of inequality. Thus, an individual’s support or opposition to the current level of economic inequality is likely to be partly determined by the level of inequality they believe exists.

Conceptual work on what constitutes ‘support’ is sparse, especially in a political or ideological context. Here, we adopt a broad definition of support laid out in previous work in political science (Easton, 1975). Easton (1975) argues that the definition of support as believing something to be “right, valid, just, or authoritative” is an adequate starting point for understanding the construct of “support” in the social sciences. Easton (1975) suggests that the key element of this definition of support is the notion of evaluation; for example, an individual’s support for economic inequality is a result of their positive evaluations of economic inequality – particularly regarding whether inequality is “right, valid, just, or authoritative.” Easton (1975) broadens this definition to contain the notion of negative evaluations. Thus, we can describe support as a person’s favorable/unfavorable or positive/negative evaluations of an object. In the simplest terms then, the extent to which a person supports economic inequality is a reflection of the extent to which they possess positive or negative evaluations of the current level of economic inequality.

Researchers often assess support for policy, such as a redistributive tax, by asking participants to provide their assessment on a dichotomous measure (e.g., support vs. oppose, vote in favor or against). However, we conceptualize support for economic inequality as more complex and continuous. For example, someone may desire that one person holds all the wealth (complete inequality), that some people hold a lot of wealth and some are poor (moderate inequality), or that every single person is equal (no inequality). Therefore, we believe that support for economic inequality exists as a continuous latent trait and is thus best measured on a spectrum.

In our definition of support for economic inequality we merge these two elements. First, we define support for economic inequality as the degree to which one possesses positive or negative evaluations towards economic inequality. Second, we often consider “support” in political domains as a direct counter to opposition. Therefore, we consider our scale to measure support for economic inequality as a range from opposition (e.g., negative or unfavorable evaluations; Easton, 1975) on the low end to support (e.g., positive or favorable) on the high end.

**Scale Development and Validation**

**Samples.** To develop a measure assessing support for economic inequality, we collected data from four separate samples on Amazon’s Mechanical Turk (mTurk). While data collection on mTurk has certain limitations (e.g., Johnson & Borden, 2012), it also affords multiple strengths. Numerous studies have demonstrated the usefulness of the mTurk population for data collection (Buhrmeister, Kwang, & Gosling, 2011; Crump, McDonnell, & Gureckis, 2013; Fort, Adda, & Cohen, 2011; Goodman, Cryder, & Cheema, 2013; Mason & Suri, 2012; Paolacci, Chandler, & Ipeirotis, 2010; Rand, 2012; Simcox & Fiez, 2014; Sprouse, 2011). These studies show, for example, that mTurk is approximately equivalent to more traditional populations (e.g., undergraduates) on dimensions such as reliability (Buhrmeister, Kwang, & Gosling, 2011; Goodman, Cryder, & Cheema, 2013), is more demographically and geographically diverse than typical university samples, and many robust psychological effects replicate on mTurk (Crump, McDonnell, & Gureckis, 2013). Moreover, mTurk is very common in the social sciences, and has been utilized effectively in recent scale development research (Armenta, Stroebe, Scheibe, Van Yperen, Stegeman, & Postmes, 2017).

**Analytic Strategy.** We followed Slaney and Maraun’s (2008) approach to data-based test analysis using an Item Response Theory (IRT) framework in which we used five steps to build and evaluate the Support for Economic Inequality scale. First, based on the definition outlined above, we generated an initial list of 18 potential items aimed at capturing the extent to which feelings about economic inequality are positive or negative. Second, we specified the formal structure of the test and identified the corresponding sense of unidimensionality – whether or not the test items measure only the one construct they were designed to measure – and the relevant statistical model with which to test unidimensionality. Third, we reduced the original 18 items to a final set of five items, removing those that performed poorly (i.e., did not adequately differentiate between people based on their support for economic inequality). Fourth, we tested the final set of five items for conformity to the relevant sense of unidimensionality. Lastly, we determined the optimal compositing rule and the resulting reliability of the final five-item scale.

In Studies 2 and 3, we repeated steps four and five using the final five-item scale with new independent samples, as well as assessed convergent and divergent validity. In addition, where Studies 1 and 2 explored support for worldwideinequality, Study 3 was conducted to test the unidimensionality and reliability of the SEIS scale referring to economic inequality in the United States. Finally, in Study 4 we explore the predictive validity of the SEIS scale by testing its association with willingness to engage in inequality mitigating action.

**Overview of Item Response Theory**

Before presenting analyses, it may be useful to give a brief non-technical overview of Item Response Theory (IRT) and define relevant terminology (for a full review see Lord, 1980; Revelle, unpublished). Traditionally, psychological tests have been constructed and evaluated using Classical Test Theory (CTT). Within the CTT framework, scores on a test are thought to represent a person’s underlying level of the trait being measured (i.e., their “true score”) and some degree of error. Assessing and improving reliability of an entire measure is the key goal when evaluating a test under a CTT framework. One of the main shortcomings with CTT, however, is that tests can only be evaluated as a whole, preventing meaningful analysis of individual items. In addition, determining how well the test functions can only be appraised with point-estimates of reliability (e.g., Cronbach’s α). Simple point-estimated reliability is limited in that it does not allow us to discern the nuances of the scale’s reliability (e.g., is the scale reliable only for people extremely high on the underlying trait, but not for people low on the trait?). IRT builds on the CTT framework by allowing analysis of how well *each individual item* behaves within the test, and allows for a broader assessment of reliability (i.e., determining how reliable the test is for different people, such as those high or low on the trait being measured). IRT accomplishes this improvement on CTT through tools such as Item Characteristic Curves, discrimination, and information functions.

For Likert scale items with more than two response options (i.e. polytomous items), Item Characteristic Curves (ICCs) provide a sense of how well each item is differentiating between people on the trait being measured by plotting latent “ability” (i.e., , the underlying level at which a person possesses the trait being measured) along the x-axis, and probability along the y-axis. Within the plot there are a series of curves, each corresponding to a response option (e.g., options 1 through 7 on a Likert scale). The height of each curve, at a given point along the x-axis, is the probability that a person at that level of will choose that option. The level of required before a response option is likely to be chosen is a *threshold.* Thus, for a 7-point item there are 6 thresholds. The point at which choosing a “2” becomes more likely than choosing a “1,” choosing “3” becomes more likely than choosing “2,” and so on. For example, in the present Study 1, item 10 demonstrates a well-defined ICC (see SOM). There are clear transition points between each response option, and possessing more extreme support for, or opposition to, inequality (as moves further from 0 in either direction), leads to increasing likelihood of selecting the more extreme response. An item that does a poor job of mapping onto the latent underlying construct will have curves that significantly overlap, with messy or disordered thresholds. For example, in the present Study 1 the ICC for item 13 (see SOM) suggests that the item is functioning dichotomously. Individuals below -1 in are most likely to choose Strongly Disagree, and individuals above -1 are most likely to choose Strongly Agree. Responding to all middle options on item 13 appears to be essentially chance, and thus uninterpretable.

Making a more detailed, numbers-driven assessment of the quality of a polytomous item involves looking at an item’s discrimination. Broadly, an item’s discrimination refers to how effectively the item can differentiate between people at a given . For example, discrimination will tell us whether the item responds to small changes in , or if large changes are required for a person to select the more extreme option. For example, does the difference between 2 and 3 on the scale reflect a small difference in latent support for economic inequality or a large difference? For polytomous items, discrimination is a function of the item’s slope and the thresholds between the response options of a given item (Muraki, 1992). Here, slope can be interpreted as the degree to which the probability of choosing a response category changes with underlying levels of (Muraki, 1992). A slope of 0 means that as changes, response choice on a Likert item does not, whereas a slope of 1 would suggest a perfect relationship: people with a more extreme choose more extreme options on a Likert item. Discrimination can, theoretically, range from 0 to infinity, with higher values (e.g., greater than 1; Bolt, Hare, Vitale, & Newman, 2004) indicating that the item does an adequate job of differentiating between people, and is thus more sensitive to changes in the underlying trait.

Building on the idea of discrimination, if we construct a test that is comprised of a set of highly discriminating items, then our test provides a great deal of “information” about the underlying latent construct and is thus more sensitive to changes in the underlying trait. Information can be thought of as an assessment of reliability across the continuum and can be viewed at both the individual item and full test level. Information is primarily displayed graphically, with along the x-axis and total information along the y-axis; a single curve plots the information over (See SOM for an example). For example, if the information curve for a test rises sharply at -1 and drops off sharply at 1 on the x-axis, this suggests the test is doing a good job measuring , and differentiating, people who fall within the -1 to 1 range of (i.e., people who are fairly moderate on the underlying trait). For people who demonstrate a high degree of the underlying trait, in this case above 1 or below -1, the test does not differentiate between them well, and is thus unreliable as a tool to measure the underlying trait in these people.

**Present Studies**

With a brief overview of IRT in mind, we shift to analyses of the Support for Economic Inequality scale. In Study 1, we employ a step-wise IRT framework (Slaney & Maraun, 2008) in which we develop an initial pool of 18 items measuring support for (worldwide) economic inequality, remove poorly functioning items, and test the resulting subset for unidimensionality according to Samejima’s Graded Response Model (Samejima, 1969). In Study 2, we confirm the results of Study 1 with a replication. In Study 3, we evaluate a United States version of the scale developed in Studies 1 and 2. Finally, in Study 4, we evaluate some initial evidence for the predictive validity of the world-wide version of the scale by exploring whether the SEIS predicts whether or not someone will sign a petition geared at addressing inequality through increasing the minimum wage.

**Study 1**

**Data**

We collected data from 604 participants (Mage = 35.6, 51.8% Female) on Amazon’s Mechanical Turk. Participants reported their agreement with the 18-item Support for Economic Inequality scale (see Table 1; items presented in random order). Afterward, participants reported their age, gender, and political orientation. We coded the items such that endorsement of higher response options indicated more positive support for economic inequality.

**Step 1: Original Item Generation**

Our goal was to begin with an expansive list of items that we could empirically narrow down to the best functioning subset. To this end, we generated an initial set of 18 items using an inductive approach (Hinkins, 1998, see Table 1 for items).

Descriptive analyses of the original 18 items revealed that most items yielded a significant positive skew, indicating that the majority of our sample endorsed response options consistent with quite strong opposition to inequality. One explanation for this skew may be that our sample consisted of individuals identifying with liberal political ideals (*M* = 5.70, SD = 2.50; with 72% of participants falling at or above the midpoint on a 1 (Conservative) to 9 (Liberal) scale), and as such, may reflect relatively extreme opposition to economic equality. Additionally, non-normality of the data is not uncommon in psychology (Woods, 2006), and the observed positive skew may indicate that, in the population, support for economic inequality demonstrates a strong positive skew such that most people feel relatively negative about economic inequality.

**Step 2: Theoretical Structure of the Test and Relevant Sense of Unidimensionality**

The Theoretical Structure (TS) of a test defines how the items were designed to measure the chosen underlying attribute, the distribution of the attribute in the population, how the test items are linked to the attribute, and whether the items are error-laden or error-free (Slaney & Maraun, 2008). That is, the statistical techniques we use to evaluate a test, and assess for unidimensionality, are inherently linked to the theoretical conceptualization of the scale. In order to determine the appropriate statistical tool with which to assess unidimensionality there are, at minimum, five components that require specification: The underlying distribution of the latent construct, the item response formats, the number of latent attributes the test is designed to measure, the form of the item/attribute regressions, and whether or not the items are error-laden (Slaney & Maraun, 2008). We chose to administer the items as 7-point polytomous items (i.e., a Likert scale ranging from 1 = Strongly Disagree to 7 = Strongly Agree) whereby people with a stronger support for, or opposition to, economic inequality should be more likely to endorse response options increasingly further above/below the midpoint, respectively. We determined the formal structure of the present Support for Economic Inequality scale to be:

|  |  |  |
| --- | --- | --- |
|  | TS{Co, 7PL, 1, 7OC, EIV} | ( 1 ) |

Thus, this is a test for which a set of 7-point Likert (7PL) items are designed as error-laden (EIV) indicators of a single underlying attribute (1; support for economic inequality), which continuously varies (Co) in degree over a population. Moreover, for any given item, the relationship between the item and the underlying attribute is conceived of as a set of seven item/attribute regressions in which the probability of endorsing any given category varies with the degree to which the individual possesses the underlying attribute (7-point ordered categorical (7OC); Slaney & Maraun, 2008). For instance, for someone who holds extremely strong support for economic inequality, the probability they will choose “7” is higher than the probability they will choose “6,” which is higher than the probability they will choose “5,” and so on.

Now that we have determined the TS, we must determine the relevant quantitative characterization of the test. The quantitative characterization describes how our test is said to behave mathematically if it is unidimensional according to our TS. When we understand how a unidimensional test should behave (i.e., the expected pattern of responses) according to our TS, we can then choose the appropriate statistical model with which to test for unidimensionality. Slaney and Maraun (2008) provide a list of common theoretical structures, their corresponding quantitative characterizations, and how to statistically test for unidimensionality in each case. Following Slaney and Maraun’s (2008) table for our specified TS, we expect our test to be unidimensional in the sense of Samejima’s Graded Response Model (GRM). Samejima’s GRM is represented as:

|  |  |  |
| --- | --- | --- |
|  |  | ( 2 ) |

This model claims that the probability of an individual with level of the underlying trait choosing a given response option, , is equal to the probability that they will choose that particular option or any lower option, (, minus the probability they will choose the option that is one higher, (; See Ostini and Nering (2006) for a more technical discussion of how these probabilities are determined and calculated. For example, the probability of someone with moderately positive support for economic inequality (e.g., 1) choosing option 5 (i.e., Slightly Agree) to a particular question is equal to the cumulative probability of them choosing options 1 through 5 minus the probability of them choosing option 6 (i.e., “Agree”).

What utilizing Samejima’s GRM means with respect to testing for unidimensionality is that we are not expecting our test to conform to the typical linear factor analytic structure; unidimensionality in our case does not mean a set of items with high factor loadings onto one latent “common factor.” Instead, the relevant test for unidimensionality is a set of quasi chi-square statistics (i.e., one for each test item) testing the null hypothesis that the observed response probabilities for each item are in line with the expected probability structure laid out by Samejima’s GRM (i.e., the probability structure laid out in Formula 2). In this case, retaining the null hypothesis suggests that the observed probabilities for each item are in line with the probabilities we would expect, and thus the set of items are unidimensional, given Samejima’s GRM.

Samejima’s GRM is the most relevant model with which to test for unidimensionality in a set of categorical Likert scale questions such as ours (as opposed to the more commonly employed linear factor analytic method; DeMars, 2010; Ostini & Nering, 2006). We will reserve our assessment of unidimensionality as part of our procedure in determining the final subset of items, as there is no need to determine that our full 18 item scale is unidimensional when we intend to delete any poorly performing items. Therefore, we now turn to our initial item reduction.

**Step 3: Initial Item Reduction/Selection**

We conducted all of our IRT individual item analyses using the ltm package (Rizopoulos, 2006) in R (R Core Team, 2016). All decisions were data-driven; we were not aware of the content of a given item during these analyses; items were labeled numerically. We based our decisions to remove items on the interaction of numerous factors; we evaluated each of the 18 items individually using their Item Characteristic Curves (ICCs), the relative proportion of information each item contributed to the scale’s overall information, and discrimination. As mentioned previously, generally higher discrimination values (e.g., above 1) are indicative of acceptable item functioning (Bolt et al., 2004).

When comparing each individual item’s ICC with its discrimination value it became clear that the five poorest functioning items were 1, 4, 6, 9, 13, and 17 (see Table 2 for discrimination values and SOM for ICCs). These six items had the lowest discrimination values, the flattest and least discerning ICCs, and contributed the lowest proportions to the scale’s overall information. Thus, these items are poor at discriminating between individuals on support for economic inequality and the response patterns are effectively random so they were dropped from further analysis.

From the remaining twelve items, we selected a small subset of five items that appeared to be the most effective: 3, 5, 8, 10, and 18. These items displayed the highest discrimination values, contributed the largest proportion of total information to the overall scale, and had the sharpest ICCs (See SOM). Given that IRT parameters, such as discrimination and information, are calculated relative to the complete set of items included in the scale, and we now had a different subset of five items, we again re-evaluated the individual item functioning for the set of five items. This analysis of the ICCs, information, and discrimination demonstrated that the set of five items function better, across the board, than the original set of 18 items (See SOM for ICCs). Discrimination values are higher, each item contributes roughly equivalent information to the scale total, and the ICCs are uniformly sharper and more cleanly defined. Following this improvement and apparent acceptable functioning of these five particular items, we turned to an assessment of unidimensionality.

**Step 4: Assessment of Unidimensionality**

To test the hypothesis of unidimensionality we utilized quasi chi-square statistics provided by the IRTPro software (Version 4.1, Scientific Software International, 2017), one for each of the five test items. These statistics quantify the difference between observed response probabilities and those expected under Samejima’s GRM (See Formula 2; Bolt et al., 2004). A large value of one or more of these statistics, relative to the degrees of freedom, constitutes evidence against the hypothesis of unidimensionality. A common problem with the chi-square approach, however, is that large sample sizes result in extremely over-powered chi-square tests that reject the null hypothesis for very small deviations from the expected probabilities, erroneously suggesting that Samejima’s GRM is a poor fit and the items are not unidimensional. Thus, we have adopted the rule under which the hypothesis of unidimensionality was rejected if any of the five quasi chi-square statistics was greater than three times its degrees of freedom (Drasgow, Levine, Tsien, Williams, & Mead, 1995).

Because all five statistics turned out to be less than three times their degrees of freedom (Table 3), the hypothesis of unidimensionality was retained (i.e., the inferential decision was made that the five items measure a common underlying trait, presumptively support for economic inequality). From here, we turn to determining the optimal compositing rule and assessing the reliability of the resulting five items.

**Step 5: Model Implied Compositing Rule and Reliability Assessment**

We computed information functions (see Figure 2a) for two of the most common candidate compositing rules for psychological scales (See https://osf.io/cmyze/ for Maple (Version 2015.1, Maplesoft 2015) worksheet containing the calculations): (1) a linear weighted estimator (using the *aj* slope parameter as the weight for each item), and (2) a unit-weighted estimator (simply adding the unweighted items together). We graphically compared the information functions of the composited scale calculated under each of these two compositing rules with the non-linear maximum likelihood estimator of (the theoretical information maximum). Figure 2a shows that neither compositing rule creates a scale that provides as much information about the underlying trait as the theoretical maximum. However, the two composited versions of the scale appear to provide identical information (the curves almost completely overlap). Given that the slope and unit weighted composited performed nearly identically, and unit weights are simpler to work with (i.e., a researcher can simply add a participant’s scores on each question, or average them, to create a single score and does not need to worry about weighting each item by its factor loading, *aj*), we decided to nominate unit weighting as the ideal compositing rule for the SEIS. Thus, future researchers who wish to use the presented SEIS should not weight the items by factor scores or any similar slope parameter, but simply compute the composite as a sum or mean for each participant. The reliability of the unit-weighted linear composite for the final five-item scale was .94 (See Maple worksheet for reliability calculations).

**Discussion**

In Study 1 we generated an initial set of 18 items measuring support for economic inequality. We distilled this set of items down to the best functioning five items (See Table 1). Finally, we determined that this set of five items appears to be unidimensional and thus measures one underlying construct, should be composited using a simple sum or mean calculation, and demonstrates a high degree of reliability. While Study 1 provides sufficient evidence for a psychometrically sound and reliable measure of support for economic inequality, we did not explore questions of convergent and divergent validity. Thus, in Study 2 we aimed to replicate the findings of Study 1 and provide additional evidence for convergent validity of the Support for Economic Inequality scale.

**Study 2**

The aim of Study 2 was to replicate and extend the findings of Study 1 with the final five-item measure. To this end, we collected data from a separate sample of adults on Amazon’s Mechanical Turk and ran all the same individual item analyses and tests of unidimensionality as Study 1. Additionally, we explored convergent validity by including measures of just world beliefs (Lambert, Burroughs, & Nguyen, 1999), support for redistribution (World Values Survey, 2014), wealth guilt (Piff, Robinson, Horberg, & Monin, unpublished), as well as face-valid measures of perceived level of inequality, perceived growth in inequality, belief that inequality is unfixable, perceived warmth and competence of people in poverty (adapted from Fiske, Cuddy, Glick & Xu, 2002), empathy (Vachon & Lynam, 2016), prosocial tendencies (Hausmann, Christiansen,& Randall, 2003), income, and political ideology. The three scales measuring perceived level and growth of inequality as well as the belief that inequality is fixable were created in-house as short, face-valid measures specifically for this project.

**Data**

We collected data from 657 participants (Mage = 23.8, 56.6% Female) on Amazon’s Mechanical Turk. Participants first completed the five-item Support for Economic Inequality scale (items presented in random order). Afterward, participants filled out, in random order, measures of just world beliefs (Lambert, Burroughs, & Nguyen, 1999), support for redistribution (World Values Survey, 2012), wealth guilt (Piff, Robinson, Horberg, & Monin, unpublished), perceived level of inequality (e.g., “Overall, the world is a fairly equal place”), perceived growth in inequality (e.g., “Economic inequality in the world is growing faster than ever before”), belief that inequality is unfixable (e.g., “Economic inequality cannot be prevented”), perceived warmth and competence of people in poverty (adapted from Fiske, Cuddy, Glick, & Xu, 2002), empathy (Vachon & Lynam, 2016), prosocial tendencies (Hausmann, Christiansen, & Randall, 2003), income, political ideology, and demographics. Identical to Study 1, we coded responses on the Support for Economic Inequality scale such that endorsement of higher response options indicated more support for economic inequality. All materials are available on https://osf.io/cmyze/.

**Individual Item Evaluation**

All individual item IRT analyses were conducted using the ltm package (Rizopoulos, 2006) in R (R Core Team, 2016). Similar to Study 1, we assessed the ICCs, information, and discrimination values for the scale in order to confirm that the items were indeed functioning appropriately with respect to their ability to discriminate between people with different underlying support for economic inequality. Replicating Study 1, all items demonstrated generally sharp and clearly defined ICCs, indicating that they function with a high degree of discrimination and are effective at differentiating between people on their underlying support for economic inequality. All discrimination values were above 3, all items contributed a high degree of information (> 9) to the total scale (Information = 62.57), and all ICCs were much clearer and discerning than the original 18 items (See SOM). These analyses confirm the selection of the final five items, which comprise a small but highly discerning and useful set of items.

**Assessment of Unidimensionality**

Consistent with Study 1, we tested the hypothesis of unidimensionality with a quasi chi-square statistic, for each of the five items, using IRTPro software (Version 4.1, Scientific Software International, 2017). Again, we also adopted the rule (Drasgow, et al., 1995) that the hypothesis of unidimensionality would be retained so long as each statistic was no greater than three times its degrees of freedom. Again, all chi-square statistics were comfortably within this range. Thus, we replicate Study 1 and retain the hypothesis that the set of five items are unidimensional, measuring one underlying trait (support for economic inequality).

**Model Implied Compositing Rule and Reliability Assessment**

We computed information functions (See Figure 2b) for the same candidate compositing rules as Study 1 (See https://osf.io/cmyze/ for Maple (Version 2015.1, Maplesoft 2015) worksheet containing calculations). Replicating Study 1, Figure 2b shows that compositing rule provided as much information as the theoretical maximum, but they again performed equally as well as each other. Given this replication of the information curves from Study 1, our decision to nominate unit weighting as the ideal compositing rule for the Support for Economic Inequality scale was confirmed. The reliability of the unit-weighted linear composite for the five-item scale was .94 (See Maple worksheet for calculations).

**Convergent and Divergent Validity**

To assess convergent and divergent validity we ran a series of correlations between the final Support for Economic Inequality scale and various measures of related political attitudes and psychological constructs. Demonstrating evidence for convergent validity, we found that more positive support for economic inequality was related to higher: general conservatism, social conservatism, economic conservatism, belief that inequality is unfixable, belief in a just world, and income. Additionally, we found that more positive support for economic inequality was related to decreased: perceived inequality, support for redistribution, wealth guilt, belief that the poor are competent, belief that the poor are warm, empathy, and prosocial tendencies (See Table 4). Demonstrating evidence for divergent validity, we found no significant relationship between support for economic inequality and gender (*r* = -.07, *p* = .07) or age (*r* = -.04, *p* = .31). Initially, we suspected there may be a correlation with age simply because older participants were likely to have higher incomes and be more conservative. However, in our sample, age was uncorrelated with both income (*r* = .03, *p* = .47) and conservatism (*r* = .02, *p* = .68).

**Discussion**

In Study 2 we replicated the findings of Study 1, demonstrating that the Support for Economic Inequality scale is comprised of a set of effective, unidimensional, reliable items. Additionally, we extended the findings of Study 1 to demonstrate evidence for convergent and divergent validity with the scale. However, Studies 1 and 2 both measure support for *worldwide* economic inequality. That is, all of the items refer to inequality *in the world today.* Therefore, in Study 3 we sought to further extend the findings of Studies 1 and 2 by testing a United States version of the Support for Economic Inequality scale.

**Study 3**

The aim of Study 3 was to extend the findings of Studies 1 and 2 and evaluate the performance of a United States version of the Support for Economic Inequality scale. The original scale developed and tested in Studies 1 and 2 was context specific, such that items assess perceptions of worldwide inequality. However, many of the researchers exploring support for economic inequality, including authors of the present study, have been particularly interested in the United States, where economic inequality is exceptionally high (Piff, Kraus, & Keltner, In Press). Therefore, we wanted to test the psychometric properties and reliability of the present scale in a high-interest context: the United States. In order to create this version we took the five-item scale and simply replaced every instance of the word “world” with “United States.” For example, “economic inequality is causing many of the world’s problems” became “economic inequality is causing many of the United States’ problems.” We then collected data from a separate sample of Amazon Mechanical Turk workers and ran all the same individual item analyses and tests of unidimensionality as in Studies 1 and 2. Additionally, we utilized the same scales to explore the convergent and divergent validity as Studies 1 and 2, as well as measures of free will, over-claiming, and socially desirable responding. Overclaiming measures self-enhancement through participants’ willingness to claim they possess knowledge they actually do not, measured by the extent to which a person claims to be knowledgeable in non-existent subjects (e.g., Plates of Parallax). We included these measures to provide evidence that the scale is not susceptible to social desirability or self-enhancement effects.

**Data**

We collected data from 619 participants (Mage = 36.01, 52.7% Female) on Amazon’s Mechanical Turk. Participants first completed the five-item U.S. version of the Support for Economic Inequality scale (items presented in random order). Following this, participants filled out, in random order, measures of just world beliefs (Lambert, Burroughs, & Nguyen, 1999), support for redistribution (World Values Survey, 2012), wealth guilt (Piff, Robinson, Horberg, & Monin, unpublished), perceived level of US inequality (e.g., “Overall, the United States is a fairly equal place”), perceived growth in US inequality (e.g., “Economic inequality in the United States is growing faster than ever before”), belief that US inequality is unfixable (e.g., “Economic inequality cannot be prevented”), warmth and competence of people in poverty (adapted from Fiske, Cuddy, Glick, & Xu, 2002), empathy (Vachon & Lynam, 2016), prosocial tendencies (Hausmann, Christiansen, & Randall, 2003), free will (Nadelhoffer, Shepard, Nahmias, Sripada, & Ross,  2014), overclaiming (Paulhus, Harms, Bruce, & Lysy, 2003), socially desirable responding (Reynolds, 1982), income, political ideology, and demographics. Similar to Studies 1 and 2, we coded the items such that endorsement of higher response options (and higher item means) indicate more positive support for economic inequality. All materials are available at https://osf.io/cmyze/.

**Individual Item Evaluation**

All individual item IRT analyses were conducted using the ltm package (Rizopolous, 2006) in R (R Core Team, 2016). We assessed the ICCs, information, and discrimination values for the scale in order to confirm that the items were indeed functioning appropriately. Consistent with Studies 1 and 2, all items demonstrated generally sharp and clearly defined ICCs, indicating that they function with a high degree of discrimination and are effective at differentiating between people on their underlying support for economic inequality. All discrimination values were above 3, all items contributed a high degree of information (> 9) to the total scale (Information = 62.76), and all ICCs were similar in profile to Study 2 (See SOM). The results of these analyses suggest that the U.S. version of our scale functions similarly to the more general worldwide version, and that the five items comprise a small but highly discerning and useful set of items.

**Assessment of Unidimensionality**

Again, following Studies 1 and 2, we computed quasi chi-square statistics for each of the five items (Table 3) using IRTPro software (Version 4.1, Scientific Software International, 2017). We adopted the same rule (Drasgow et al., 1995) such that the hypothesis of unidimensionality would be retained so long as each statistic was no greater than three times its degrees of freedom. Replicating Studies 1 and 2, all quasi chi-square statistics were within this range. Thus, we again retain the hypothesis that the five items are unidimensional and measure one underlying trait (support for economic inequality).

**Model Implied Compositing Rule and Reliability Assessment**

We computed information functions (see Figure 2c) for the same candidate compositing rules as Studies 1 and 2 (See https://osf.io/cmyze/ for Maple (Version 2015.1, Maplesoft 2015) worksheet containing calculations). Replicating Studies 1 and 2, this analysis demonstrated that neither linear estimator performed as well as the theoretical maximum, and the unit weighted composite performed only slightly worse than the *aj*weighted composite. Given this replication of the information curves from Studies 1 and 2, our decision was to again nominate unit weighting as the ideal compositing rule for the United States version of the SEIS. While the unit-weighted version appears to give a little bit less information, we do not believe the loss of information is great enough to justify significantly complicating the compositing rule. The reliability of the unit-weighted linear composite for the final five-item United States version of the scale was .94 (See Maple worksheet for calculations).

**Convergent and Divergent Validity**

To assess convergent and divergent validity we ran a series of correlations between the American version of the SEIS and the same measures of related political attitudes and psychological constructs as Study 2, plus one additional measure of free will. Similar to Study 2, as evidence for convergent validitywe found that more support for economic inequality was related to higher: general conservatism, social conservatism, economic conservatism, belief that inequality is unfixable, belief in a just world, free will, and income (see Table 5). Additionally, we found that more support for economic inequality was related to decreased: perceived inequality, support for redistribution, wealth guilt, belief that the poor are competent, belief that the poor are warm, empathy, and prosocial tendencies.Additionally, as evidence of divergent validity, we found no relationship between support for economic inequality and gender (*r* = -.06, *p* = .16) or age (*r* = -.03, *p* = .51). Support for economic inequality was uncorrelated with both overclaiming (*r* = -.04, *p* = .38) and socially desirable responding (*r* = .00, *p* = .98).

**Context Specific Scale Differences.** In order to explore if mean responses differ due to the question’s wording we explored whether or not there was a significant mean difference between the current United States and worldwide versions of the scale (see OSF page for all analyses). To do so, we combined the data from Study 2 and Study 3 into one data file containing responses to the five SEIS items with framing coded as 1 (world framing) and 2 (U.S. framing). We conducted an independent-samples t-test comparing the mean composited scale and found no significant difference (*M*world = 2.71, *SD* = 1.43; *M*USA = 2.72, *SD* = 1.46; *t*(1221) = -.12, *p* = .90, 95% CI [-.17,.15], *d* = 0.01). Further, in looking at the measures of dispersion across the world and U.S. specific mean composited scales, we see similar measures of skewness (0.80 and 0.89, respectively) and kurtosis (0.02 and 0.25, respectively). As such, the two versions of the SEIS are roughly equivalent in terms of their descriptive statistics. This suggests (if adjudicated appropriately) the SEIS can be flexibly adapted to suit a researcher’s needs; for example, the scale can be altered to measure support for economic inequality in different contexts such as a specific country, state, county, or specific city.

One final piece of evidence that is crucial in demonstrating a useful scale is predictive validity. Specifically, does the SEIS predict relevant behavior? We tested this question in Study 4 by exploring whether support for inequality predicts the signing of a minimum wage petition.

**Study 4**

**Participants**

We collected data from 117 participants (Mage = 35.31, 59.0% Female) on Amazon’s Mechanical Turk.[[1]](#endnote-1)

**Procedure**

Participants were first presented with a petition for a 70% increase in the minimum wage in the United States from $7.25/hour to $10.10/hour (See OSF for full petition content). Following reading the petition, participants were asked “to what extent do you agree with the content of the petition” on a 1 (Strongly disagree) to 7 (Strongly agree) scale, and were subsequently given the opportunity to sign the petition by entering their first and last name, email address, and zip code. After choosing whether or not to sign the petition, participants filled out the worldwide version of the SEIS and demographics (age, gender, education level, household income, and the level of inequality present in the area in which they grew up; see OSF for full questionnaire).

If the SEIS has adequate predictive validity, higher support for economic inequality should predict (a) decreased agreement with the content of the petition and (b) decreased likelihood of signing the petition. We predicted these relationships would hold controlling for age, gender, education, household income, political ideology, and the level of inequality that one experienced in their childhood neighborhood.

**Results**

First, a linear regression of agreement with the content of the petition on support for economic inequality confirmed our hypothesis, such that higher support for economic inequality was associated with lower reported agreement with the petition (*b* = -.67, *p* < .001). Importantly, this relationship held controlling for all of the relevant included variables that might influence one’s agreement with a petition aimed at increasing minimum wage (*b* = -.61, *p* < .001; See Table 6).

Second, a logistic regression found that higher support for economic inequality was associated with lower likelihood of signing the petition (*b* = -.44, *p* = .002), further confirming our hypothesis. Importantly, this relationship held controlling for relevant variables (*b* = -.51, *p* = .02; See Table 7).

**Discussion**

The results of Study 4 show predictive validity of the SEIS. As support for economic inequality increases, agreement with the content of the petition decreases. More important, as support for economic inequality increases, the probability that one will actually sign the petition goes down. Both of these relationships remain after controlling for age, gender, education, household income, political ideology, and level of inequality in a participant’s childhood neighborhood. Additionally, in this model, support for economic inequality was the strongest predictor of both agreement with the content of the petition as well as signing of the petition compared to all other entered covariates. These findings bolster the predictive validity of the SEIS, indicating that it captures people’s feelings toward economic inequality and predicts their willingness to take action to mitigate it.

**General Discussion**

As economic inequality has risen in countries like the United States, so, too, has its status as a topic of interest to researchers, politicians, policy makers, and the public. However, there are presently no psychometrically adjudicated and validated measures of support for economic inequality. The present paper seeks to fill this gap by providing a psychometrically sound scale of support for economic inequality. Across four studies we take an Item Response Theory (i.e., Slaney & Maraun, 2008) approach to construct, evaluate, and validate two short scales measuring support for economic inequality in the world and U.S. contexts. The final five-item scales demonstrate favorable psychometric properties (individual item functioning and unidimensionality), high reliability, predictive validity, and convergent and divergent validity for both the worldwide and United States contexts.

Using this measure, future researchers have an efficient and effective tool for measuring support for economic inequality. As mentioned previously, most research studying related constructs (e.g., general perceptions of inequality) have relied on either face valid measures, or previously collected data, such as the International Social Survey Programme and the World Values Survey. While these data are extremely valuable in uncovering relationships at the national or international scale, it is our hope that the present measures will aid researchers in understanding how people develop support for economic inequality, as well as evaluating the downstream consequences of this attitude. For example, this measure could be used to assess the effectiveness of interventions designed to reduce support for economic inequality, or understand the causes of support for economic inequality.

**Limitations and Future Directions**

The presented studies provide a new and effective measure of support for economic inequality, but there are numerous aspects of the scale still to be explored. For instance, it would be worthwhile to test the scale for differential item functioning in various populations. While our samples generally contained a wide spread of ages, incomes, and ideologies, we did not assess whether the scale performed equally in these populations due to sample size constraints. It is possible that when exploring the individual item functioning, the items of our final scale would capture different ranges of the underlying trait in, for example, conservatives, higher income people, or different racial groups.

One additional limitation is the overall skewness of the responses to the scale. However, non-normality of the data is not uncommon in psychology (Woods, 2006), and the observed positive skew may indicate that most people in the population have low levels of support for economic inequality. Consistent with this explanation, in a large cross-national sample of Europeans (n = 54,059) the mean response to the question “Income differentials in [my] country are too large” is 4.23 on a 5-point Likert scale, where 5 is strongly agree (skewness = -1.34, kurtosis = 1.70; ISSP, 2009). This large-scale cross-national sample suggests that, in general, people, here a large sample of Europeans, are quite strongly intolerant of inequality. As such, we do not interpret the skewness in the data to be problematic, but rather a reflection of how this construct manifests at the population level.

One major strength of the present scale is its adaptability. Specifically, researchers can modify the measure to different levels and contexts with ease. For example, the scale can easily be adapted to assess support for economic inequality in any other country, a specific state, city, county, or community, etc. While we only explored one of these possible iterations – the United States – we encourage researchers to use the scale in the context that works for their research question, but not without proper adjudication of the scale in each new form. It is not to be assumed that the items would function akin to how they functioned in the present paper when their core content has been changed, though the results of Study 3 suggests that the scale can be easily adapted to different populations.

Through future utilization of the SEIS, we can explore the downstream consequences of support for economic inequality and build an understanding of the social and psychological factors that influence the development of one’s (non)support for economic inequality. For example, how does the county- or individual-level social mobility where one grows up influence her level of support for economic inequality? Moreover, future researchers interested in support for economic inequality can use the SEIS to explore how it relates to non-self-report responses to economic inequality (e.g., physiological arousal). Finally, as researchers use the SEIS across different contexts, we only stand to further strengthen evidence for its psychometric fitness, reliability, and validity.

References

Armenta, B. M., Stroebe, K., Sheibe, S., Van Yperen, N. W., Stegeman, A., & Postmes, T. (2017). Permeability of group boundaries: Development of the concept and a scale. *Personality and Social Psychology Bulletin, 43*, 418-433. doi:10.1177/0146167216688202

Attanasio, O. P., & Pistaferri, L. (2016). Consumption inequality. *Journal of Economic Perspectives, 30*, 3-28. doi:10.1257/jep.30.2.3

Blomquist, N. S. (1981). A comparison of distributions of annual and lifetime income: Sweden around 1970. *The review and income and wealth, 27*, 243-264. doi:10.1111/j.1475-4991.1981.tb00227.x

Buhrmeister, M., Kwang, T., & Gosling, S. D. (2011). Amazon’s mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3-5. doi:10.1177/1745691610393980

Buttrick, N. R., & Oishi, S. (2017). The psychological consequences of income inequality. *Social and Personality Psychology Compass, 11*, e12304. doi:10.1111/spc3.12304

Crick, B. (1971). Toleration and tolerance in theory and practice. *Government and Opposition, 6*, 44-171.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon’s mechanical turk as a tool for experimental behavioral research. *PLoS ONE.* doi:10.1371/journal.pone.0057410

DeMars, C. (2010). *Item Response Theory.* New York, NY: Oxford.

Easton, D. (1975). A re-assessment of the concept of political support. *British Journal of Political Science, 5*, 435-457. doi:10.1017/s0007123400008309

Engelen, B., & Nys, T. (2008). Tolerance: A virtue?: Towards a broad and descriptive definition of tolerance. *Philosophy in the Contemporary World, 15*(1), 44-54.

Engelhardt, C., & Wagener, A. (2014). Biased perceptions of income inequality and redistribution. *CESIfo Working Paper Series No. 4838.* Retrieved from: https://ssrn.com/abstract=2463129

Fort, K., Addy, G., & Cohen, K. B. (2011). Amazon mechanical turk: Gold mine or coal mine? *Computational Linguistics, 37*, 413-420. doi:10.1162/COLI\_a\_00057

Gallup. (2016). *Economy continues to rank as top U.S. problem.* Retrieved from http://www.gallup.com/poll/191513/economy-continues-rank-top-problem.aspx

Gilbert, D. T., & Gill, M. J. (2000). The momentary realist. *Psychological Science, 11*, 394-398. doi:10.1111/1467-9280.00276

Gilovich, T. (1990). Differential construal and the false consensus effect. *Journal of Personality and Social Psychology, 59*, 623-634. doi:10.1037/0022-3514.59.4.623

Gini, C. (1912). *Variabilità e mutabilità.* Reprinted in Pizetti, E.; Salvemini, T.(Eds.),(1955). *Memorie di metodologica statistica*. Rome: Libreria Eredi Virgilio Veschi.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2012). Data collection in a flat world: The strengths and weaknesses of mechanical turk samples. *Behavioral Decision Making, 26*, 213-224. doi:10.1002/bdm.1753

Hillen, M. A., Gutheil, C. M., Strout, T. D., Smets, E. M. A., Han, P. K. J. (2017). Tolerance of uncertainty: Conceptual analysis, integrative model, and implications for healthcare. *Social Science and Medicine, 180*, 62-75. doi:10.1016/j.socscimed.2017.03.024

Hine, D. W., & Montiel, C. J. (1999). Poverty in developing nations: A cross-cultural attributional analysis. *European Journal of Social Psychology, 29*, 943-959. doi: 10.1002/(SICI)1099-0992(199911)29:7<943::AID-EJSP978>3.0.CO;2-5

Horton, J. (1996). Toleration as a Virtue. In D. Heyd (Ed.), *Toleration: An elusive virtue* (pp. 28-43). Princeton, NJ: Princeton University Press.

Johnson, D. R., & Borden, L. A. (2012). Participants at your fingertips: Using Amazon’s mechanical turk to increase student-faculty collaborative research. *Teaching of Psychology, 39*, 245-251. doi:10.1177/0098628312456615

King, P. (1971). The problem of tolerance. *Government and Opposition, 6*, 172-207.

Kraus, M. W., Park, J. W., & Tan, J. J. X. (2017). Signs of social class: The experience of economic inequality in everyday life. *Perspectives on Psychological Science, 12*, 422-435. doi:10.1177/1745691616673192

Kraus, M. W., Piff, P. K., & Keltner, D. (2009). Social class, sense of control, and social explanation. *Journal of Personality and Social Psychology, 97*, 992-1004. doi:10.1037/a0016357

Lord, F. M. (1980). *Application of Item Response Theory to Practical Testing Problems.* Hillsdale, NJ: Erlbaum.

Mason, W., & Suri. S. (2012). Conducting behavioral research on Amazon’s mechanical turk. *Behavioral Research Methods, 44*, 1-23. doi:10.3758/s13428-011-0124-6

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *ETS Research Report Series, 1992,* 1-30. doi:10.1002/j.2333-8504.1992.tb01436.x

Murphy, C. N. (2001). Political consequences of new inequality. *International Studies Quarterly, 45*, 347-356.

Norton, M. I., & Ariely, D. (2011). Building a better America – One wealth quintile at a time. *Perspectives on Psychological Science, 6*, 9-12. doi:10.1177/1745691610393524

Novotny, J. (2007). On the measurement of regional inequality: Does spatial dimension of income inequality matter? *Annals of Regional Science, 41*, 563-580.

doi:10.1007/s00168-007-0113-y

Obama, B. H. (2013, December 4). *Remarks by the President on Economic Mobility.* Retrieved from https://obamawhitehouse.archives.gov/the-press-office/2013/12/04/remarks-president-economic-mobility

Osberg, L., & Smeeding, T. (2006). “Fair” inequality? Support for pay differentials: The United States in comparative perspective. *American Sociological Review, 71*(3), 450-473.

Ostini, R., & Nering, M. L. (2006). *Quantitative Applications in the Social Sciences: Polytomous Item Response Theory Models.* Thousand Oaks, CA: SAGE Publications, Inc

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon’s mechanical turk. *Judgment and Decision Making, 5*, 411-419. doi:

Pickett, K. E., & Wilkinson, R. G. (2017). Immorality of inaction on inequality. *British Medical Journal, 356*. doi:10.1136/bmj.j556

Piff, P. K., Robinson, A. R., Horberg, E. J., Monin, B. unpublished manuscript. An embarrassment of riches: The expression of wealth guilt in social interactions.

Piff, P. K., Kraus, M. W., & Keltner, D. (In Press). Unpacking the inequality paradox: The psychological roots of inequality and social class. *Advances in Experimental Social Psychology.* doi:10.1016/bs.aesp.2017.10.002

Piketty, T., & Saez, E. (2014). Inequality in the long run. *Science, 344*, 838-844. doi:10.1126/science.1251936

Rand, D. G. (2012). The promise of mechanical turk: How online labor markets can help theorists behavioral experiments. *Journal of Theoretical Biology, 299*, 172-179. doi:10.1026/j.jtbi.2011.03.004

Revelle, W. *An Introduction to Psychometric Theory with Applications in R.* Retrieved from http://www.personality-project.org/r/book/

Samejima, F. (1969). *Estimation of Latent Ability Using a Response Pattern of Graded Scores.* (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society. Retrieved from https://www.psychometricsociety.org/sites/default/files/pdf/MN17.pdf

Savani, K., & Rattan, A. (2012). A choice mind-set increases the acceptance and maintenance of wealth inequality. *Psychological Science, 23*, 796-804. doi:10.1177/0956797611434540

Shariff, A. F., Wiwad, D., & Aknin, L. B. (2016). Income mobility breeds tolerance for income inequality: Cross-national and experimental evidence. *Perspectives on Psychological Science, 11*, 373-380. doi:10.1177/1745691616635596

Simcox, T., & Fiez, J. A. (2014). Collecting response times using Amazon’s mechanical turk and Adobe flash. *Behavioral Research Methods, 46*, 95-111. doi:10.3758/s13428-013-0345-y

Slaney, K. L., & Maraun, M. D. (2008). A proposed framework for conducting data-based test analysis. *Psychological Methods, 13*, 376-390. doi:10.1037/a0014269

Sprouse, J. (2011). A validation of Amazon mechanical turk for the collection of acceptability judgments in linguistic theory. *Behavioral Research Methods, 43*, 155-167. doi:10.3758/s13428-010-0039-7

Theil, H. (1967). *Economics and Information Theory.* Amsterdam: North-Holland.

Wilkinson, R. G. (2006). *The Impact of Inequality: How to Make Sick Societies Healthier.* New York, NY: The New Press.

Wolff, E. N. (2010). *Recent trends in household wealth in the United States: Rising debt and the middle-class squeeze—An update to 2007*. (Working Paper No. 598). Retrieved from the Levy Economics Institute Working Paper Collection Website http://www.levyinstitute.org/pubs/wp\_589.pdf

Woods, C. M. (2006). Ramsay-curve item response theory (rc-irt) to detect and correct for nonnormal latent variables. *Psychological Methods, 11*, 253-270. doi:10.1037/1082-989X.11.3.253

Yitzhaki, S. (1998). More than a dozen alternative ways of spelling Gini. *Research on Economic Inequality, 8*, 13-30.

Table 1. Original 18 items, with their descriptive statistics from Study 1. **(R)** denotes item is reverse coded. Descriptive statistics were calculated after each relevant item was reverse scored.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Item content | Mean | SD | Skewness | Kurtosis |
| 1. | Economic inequality is one of the biggest problems in today’s world. **(R)** | 2.20 | 1.84 | 0.56 | -0.82 |
| 2. | Economic inequality is not a big problem in the world | 1.51 | 1.65 | 1.00 | -0.05 |
| **3.** | **The negative consequences of economic inequality have been largely exaggerated.** | **1.66** | **1.67** | **0.91** | **-0.10** |
| 4. | Economic inequality is mostly caused by different levels of individual effort. | 2.24 | 1.72 | 0.30 | -0.96 |
| **5.** | **Economic inequality is causing many of the world’s problems. (R)** | **1.90** | **1.64** | **0.76** | **-0.18** |
| 6. | Economic inequality is inherently unfair. **(R)** | 1.96 | 1.72 | 0.70 | -0.41 |
| 7. | I am not bothered by the current level of economic inequality in the world. | 1.77 | 1.83 | 0.87 | -0.36 |
| **8.** | **I am very disturbed by the amount of economic inequality in the world today. (R)** | **1.87** | **1.74** | **0.77** | **-0.37** |
| 9. | There are much bigger problems in the world than economic inequality. | 3.02 | 1.82 | -0.11 | -0.98 |
| **10.** | **Economic inequality is not a problem.** | **1.25** | **1.50** | **1.20** | **0.60** |
| 11. | There are some positive benefits that result from economic inequality. | 1.34 | 1.49 | 0.98 | -0.05 |
| 12. | Overall, economic inequality is good for the world. | 1.44 | 1.55 | 0.98 | 0.12 |
| 13. | I wish there was more economic inequality in the world. | 1.23 | 1.71 | 1.40 | 0.93 |
| 14. | Economic inequality is fair. | 1.63 | 1.70 | 0.91 | -0.12 |
| 15. | I am very concerned about the current level of economic inequality in the world. **(R)** | 1.93 | 1.72 | 0.78 | -0.27 |
| 16. | If I could, I would make the world a more equal place. **(R)** | 1.44 | 1.49 | 1.12 | 0.89 |
| 17. | Economic inequality does not lead to anything good. | 2.14 | 1.72 | 0.33 | -1.02 |
| **18.** | **We need to do everything possible to reduce economic inequality in the world today. (R)** | **1.87** | **1.64** | **0.81** | **-0.04** |

*Note.* The bolded items are the final five-item “Support for Economic Inequality” scale, as determined in Study 1.

Table 2. Graded model parameter estimates for the full 18 items in Study 1.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Item | Discrimination (*a*) | Threshold 1 (*b*1) | Threshold 2 (*b*2) | Threshold 3 (*b*3) | Threshold 4 (*b*4) | Threshold 5 (*b*5) | Threshold 6 (*b*6) |
| 1 | 2.73 | -1.79 (.10) | -1.17 (.07) | -0.81 (.06) | -0.49 (.05) | 0.14 (.06) | 0.94 (.08) |
| 2 | 3.57 | -2.05 (.11) | -1.59 (.08) | -1.17 (.06) | -0.91 (.05) | -0.38 (.05) | 0.31 (.06) |
| 3 | 3.79 | -1.96 (.11) | -1.47 (.07) | -1.12 (.06) | -0.76 (.05) | -0.28 (.05) | 0.49 (.06) |
| 4 | 1.67 | -2.73 (.20) | -1.78 (.12) | -0.97 (.08) | -0.31 (.07) | 0.28 (.07) | 1.16 (.10) |
| 5 | 3.03 | -2.01 (.11) | -1.45 (.07) | -1.07 (.06) | -0.69 (.05) | 0.06 (.05) | 0.83 (.07) |
| 6 | 2.48 | -2.02 (.12) | -1.54 (.08) | -1.07 (.07) | -0.66 (.06) | 0.03 (.06) | 0.80 (.07) |
| 7 | 2.94 | -1.86 (.10) | -1.43 (.08) | -1.05 (.06) | -0.81 (.06) | -0.26 (.05) | 0.43 (.06) |
| 8 | 3.87 | -1.78 (.09) | -1.32 (.07) | -0.96 (.05) | -0.66 (.05) | -0.06 (.05) | 0.64 (.06) |
| 9 | 1.69 | -1.85 (.12) | -1.11 (.08) | -0.27 (.07) | 0.38 (.07) | 0.94 (.09) | 1.66 (.12) |
| 10 | 4.28 | -2.25 (.13) | -1.73 (.09) | -1.30 (.06) | -0.98 (.05) | -0.54 (.05) | 0.15 (.05) |
| 11 | 3.19 | -2.87 (.24) | -1.86 (.10) | -1.38 (.07) | -0.93 (.06) | -0.52 (.05) | 0.27 (.06) |
| 12 | 3.15 | -2.21 (.13) | -1.83 (.10) | -1.32 (.07) | -0.89 (.06) | -0.42 (.05) | 0.33 (.06) |
| 13 | 1.30 | -2.93 (.25) | -2.44 (.20) | -1.98 (.16) | -1.48 (.13) | -1.05 (.10) | -0.14 (.08) |
| 14 | 2.39 | -2.20 (.14) | -1.75 (.10) | -1.32 (.08) | -0.84 (.06) | -0.33 (.06) | 0.45 (.06) |
| 15 | 3.50 | -1.74 (.09) | -1.38 (.07) | -0.96 (.06) | -0.65 (.05) | -0.02 (.05) | 0.76 (.07) |
| 16 | 3.08 | -2.14 (.12) | -1.73 (.09) | -1.46 (.08) | -0.95 (.06) | -0.29 (.05) | 0.43 (.06) |
| 17 | 2.01 | -2.59 (.18) | -1.70 (.10) | -0.92 (.07) | -0.37 (.06) | 0.12 (.06) | 0.95 (.08) |
| 18 | 3.13 | -1.89 (.10) | -1.52 (.08) | -1.11 (.06) | -0.66 (.05) | -0.03 (.05) | 0.80 (.07) |

*Note.* Standard Errors for each parameter are in brackets. *a* is the item’s discrimination parameter, *b*s are the thresholds.

Table 3. Goodness-of-fit Chi-Square tests for the five item scale in Studies 1, 2, and 3.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Study 1 | | | | Study 2 | | | | Study 3 | | | |
| Item | Chi-square | df | p-value | Chi-square/df | Chi-Square | df | p-value | Chi-square/df | Chi-square | df | p-value | Chi-square/df |
| 3 | 84.37 | 69 | .10 | 1.22 | 142.62 | 75 | < .001 | 1.90 | 108.18 | 76 | .009 | 1.42 |
| 5 | 84.91 | 71 | .12 | 1.20 | 108.28 | 71 | .003 | 1.53 | 98.73 | 62 | .002 | 1.59 |
| 8 | 74.61 | 64 | .17 | 1.17 | 132.38 | 71 | < .001 | 1.86 | 136.23 | 63 | < .001 | 2.16 |
| 10 | 89.32 | 69 | .05 | 1.29 | 159.82 | 82 | < .001 | 1.95 | 133.31 | 75 | < .001 | 1.78 |
| 18 | 95.29 | 71 | .03 | 1.34 | 129.45 | 72 | < .001 | 1.80 | 113.83 | 67 | < .001 | 1.70 |

Table 4. Correlations between all scales assessing convergent validity in Study 2.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *1.* | *2.* | *3.* | *4.* | *5.* | *6.* | *7.* | *8.* | *9.* | *10.* | *11.* | *12.* | *13.* |
| 1. SEIS | -- |  |  |  |  |  |  |  |  |  |  |  |  |
| 2. GC | .58\*\*\* | -- |  |  |  |  |  |  |  |  |  |  |  |
| 3. SC | .52\*\*\* | .89\*\*\* | -- |  |  |  |  |  |  |  |  |  |  |
| 4. EC | .57\*\*\* | .87\*\*\* | .75\*\*\* | -- |  |  |  |  |  |  |  |  |  |
| 5. IU | .72\*\*\* | .49\*\*\* | .45\*\*\* | .51\*\*\* | -- |  |  |  |  |  |  |  |  |
| 6. PI | .56\*\*\* | .32\*\*\* | .31\*\*\* | .30\*\*\* | .40\*\*\* | -- |  |  |  |  |  |  |  |
| 7. SR | -.78\*\*\* | -.52\*\*\* | -.47\*\*\* | -.54\*\*\* | -.68\*\*\* | -.41\*\*\* | -- |  |  |  |  |  |  |
| 8. WG | -.26\*\*\* | -.21\*\*\* | -.18\*\*\* | -.25\*\*\* | -.18\*\*\* | .05 | .26\*\*\* | -- |  |  |  |  |  |
| 9. Comp | -.23\*\*\* | -.11\*\* | -.08 | -.18\*\*\* | -.21\*\*\* | .02 | .22\*\*\* | .26\*\*\* | -- |  |  |  |  |
| 10. Warm | -.25\*\*\* | -.07 | -.08\* | -.14\*\*\* | -.24\*\*\* | -.03 | .24\*\*\* | .18\*\*\* | .71\*\*\* | -- |  |  |  |
| 11. Emp | -.28\*\*\* | -.11\*\* | -.12\*\* | -.12\*\* | -.25\*\*\* | -.39\*\*\* | .24\*\*\* | -.04 | -.03 | .13\*\*\* | -- |  |  |
| 12. PT | -.20\*\*\* | -.05 | -.06 | -.10\* | -.16\*\*\* | -.10\*\* | .17\*\*\* | .19\*\*\* | .23\*\*\* | .30\*\*\* | .40\*\*\* | -- |  |
| 13. BJW | .33\*\*\* | .29\*\*\* | .29\*\*\* | .28\*\*\* | .31\*\*\* | .36\*\*\* | -.26\*\*\* | .01 | -.01 | .00 | .01 | .08\* | -- |
| 14. Inc | .11\*\* | .09\* | .06 | .13\*\*\* | .16\*\*\* | .01 | -.09\* | .07 | -.15\*\*\* | -.16\*\*\* | .02 | .03 | .09\* |

*Note.* SEIS = Support for Economic Inequality; GC = General Conservatism; SC = Social Conservatism; EC = Economic Conservatism; IU = Belief that Inequality is Unfixable; PI = Perceived Inequality; SR = Support for Redistribution; WG = Wealth Guilt; Comp = Perceptions of the poor as competent; Warm = Perceptions of the poor as warm; Emp = Empathy; PT = Prosocial Tendencies; BJW = Belief in a Just World; Inc = Income. \* = *p* < .05, \*\* = *p* < .01, \*\*\* = *p* < .001

Table 5. Correlations between all scales assessing convergent validity in Study 3.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *1.* | *2.* | *3.* | *4.* | *5.* | *6.* | *7.* | *8.* | *9.* | *10.* | *11.* | *12.* | *13.* | *14.* |
| 1. SEIS | -- |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2. GC | .59\*\*\* | -- |  |  |  |  |  |  |  |  |  |  |  |  |
| 3. SC | .53\*\*\* | .89\*\*\* | -- |  |  |  |  |  |  |  |  |  |  |  |
| 4. EC | .58\*\*\* | .88\*\*\* | .78\*\*\* | -- |  |  |  |  |  |  |  |  |  |  |
| 5. IU | .67\*\*\* | .51\*\*\* | .47\*\*\* | .50\*\*\* | -- |  |  |  |  |  |  |  |  |  |
| 6. PI | .70\*\*\* | .44\*\*\* | .40\*\*\* | .46\*\*\* | .51\*\*\* | -- |  |  |  |  |  |  |  |  |
| 7. SR | -.73\*\*\* | -.52\*\*\* | -.44\*\*\* | -.54\*\*\* | -.63\*\*\* | -.51\*\*\* | -- |  |  |  |  |  |  |  |
| 8. WG | -.18\*\*\* | -.18\*\*\* | -.15\*\*\* | -.19\*\*\* | -.13\*\*\* | .02 | .20\*\*\* | -- |  |  |  |  |  |  |
| 9. Comp | -.25\*\*\* | -.20\*\*\* | -.18\*\*\* | -.18\*\*\* | -.20\*\*\* | -.01\*\* | .27\*\*\* | .30\*\*\* | -- |  |  |  |  |  |
| 10. Warm | -.31\*\*\* | -.21\*\*\* | -.21\*\*\* | -.20\*\*\* | -.25\*\*\* | -.15\*\*\* | .32\*\*\* | .24\*\*\* | .73\*\*\* | -- |  |  |  |  |
| 11. Emp | -.27\*\*\* | -.10\* | -.21\* | -.11\*\* | -.28\*\*\* | -.32\*\*\* | .24\*\*\* | -.09\* | -.01 | .19\*\*\* | -- |  |  |  |
| 12. PT | -.16\*\*\* | -.06 | -.06 | -.05 | -.14\*\*\* | -.04 | .20\*\*\* | .24\*\*\* | .23\*\*\* | .28\*\*\* | .38\*\*\* | -- |  |  |
| 13. BJW | .30\*\*\* | .29\*\*\* | .27\*\*\* | .30\*\*\* | .30\*\*\* | .40\*\*\* | -.23\*\*\* | .05 | .00 | -.01 | -.01 | .15\*\*\* | -- |  |
| 14. FW | .27\*\*\* | .27\*\*\* | .25\*\*\* | .24\*\*\* | .26\*\*\* | .28\*\*\* | -.23\*\*\* | -.14\*\*\* | -.13\*\*\* | -.09\* | .07 | .09\*\* | .51\*\*\* | -- |
| 15. Inc | .15\*\*\* | .08\* | .06 | .16\*\*\* | .13\*\*\* | .11\*\* | -.13\*\* | .06 | -.14\*\*\* | -.11\*\* | .03 | .03 | .20\*\*\* | .11\*\* |

*Note.* SEIS = Support for Economic Inequality; GC = General Conservatism; SC = Social Conservatism; EC = Economic Conservatism; IU = Belief that Inequality is Unfixable; PI = Perceived Inequality; SR = Support for Redistribution; WG = Wealth Guilt; Comp = Perceptions of the poor as competent; Warm = Perceptions of the poor as warm; Emp = Empathy; PT = Prosocial Tendencies; BJW = Belief in a Just World; FW = Free Will; Inc = Income. \* = *p* < .05, \*\* = *p* < .01, \*\*\* = *p* < .001

Table 6. Linear regressions of agreement with the content of the petition onto the SEIS scale, controlling for relevant covariates.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model 1:  SEIS only | | Model 2: SEIS plus  covariates | |
|  | *b* | SE | *b* | SE |
| Intercept | 7.29\*\*\* | 0.25 | 5.83\*\*\* | 0.86 |
| SEIS | -0.68\*\*\* | 0.07 | -0.61\*\*\* | 0.09 |
| Age | -- | -- | -0.01 | 0.01 |
| Gender | -- | -- | 0.38 | 0.23 |
| Education | -- | -- | 0.42\* | 0.18 |
| Household Income | -- | -- | 0.02 | 0.06 |
| Political Ideology | -- | -- | -0.08 | 0.08 |
| Childhood Inequality | -- | -- | 0.06 | 0.08 |
| **Adjusted R2** | .41 | | .45 | |

*Note. \* p* < .05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

Table 6. Logistic regressions of whether or not the petition was signed onto the SEIS scale, controlling for relevant covariates.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model 1:  SEIS only | | Model 2: SEIS plus  covariates | |
|  | *b* | SE | *b* | SE |
| Intercept | 0.91\*\*\* | 0.41 | -5.26\*\* | 1.82 |
| SEIS | -0.45\*\* | 0.15 | -0.51\* | 0.22 |
| Age | -- | -- | -0.01 | 0.02 |
| Gender | -- | -- | 0.99\* | 0.45 |
| Education | -- | -- | 0.88\* | 0.37 |
| Household Income | -- | -- | 0.19 | 0.13 |
| Political Ideology | -- | -- | -0.04 | 0.17 |
| Childhood Inequality | -- | -- | 0.48\*\* | 0.18 |

*Note.* For petition signing, 0 = did not sign, 1 = did sign. *\* p* < .05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

Figure 1. The number of published articles regarding “income inequality” or “economic inequality” indexed on Web of Science between 1993 and 2016.

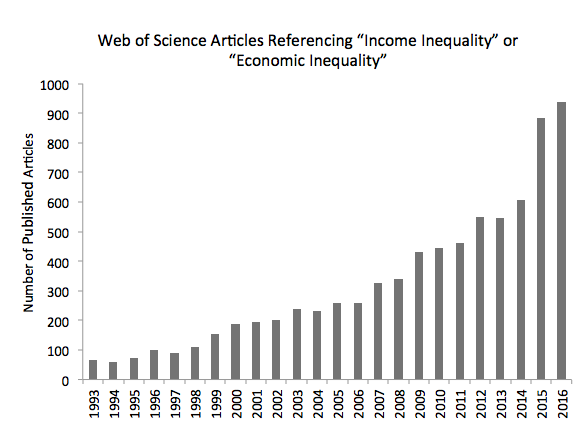


Figure 2a. Study 1 Information Function for the three candidate compositing rules. The solid line is the Maximum Likelihood estimated theoretical maximum information, the dashed line is the *aj* weighted composite information, and the dotted line is the unit-weighted composite information.

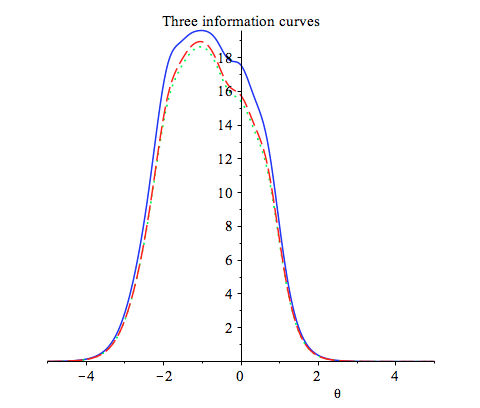
**

Figure 2b. Study 2 Information Function for the three candidate compositing rules. The solid line is the Maximum Likelihood estimated theoretical maximum information, the dashed line is the *aj* weighted composite information, and the dotted line is the unit-weighted composite information.

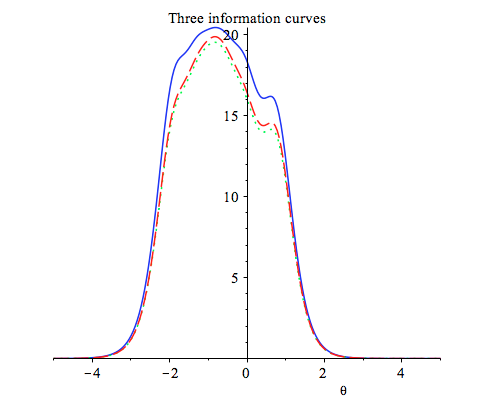
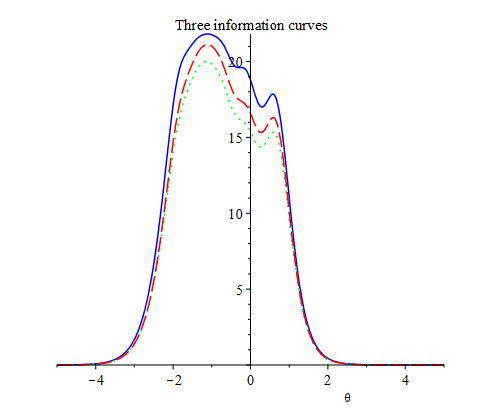


Figure 2c. Study 3 Information Function for the three candidate compositing rules. The solid line is the Maximum Likelihood estimated theoretical maximum information, the dashed line is the *aj* weighted composite information, and the dotted line is the unit-weighted composite information.



1. These data were part of a larger study in which there was a manipulation. Participants either read an article about (a) hardworking poor people in America, (b) about the poor in America more broadly, or (c) about gun violence in America. These articles can be found on the OSF page for this paper. Condition assignment did not impact scores on the SEIS, F(1,115) = 1.012, p = .32. [↑](#endnote-ref-1)