Accepted for Publication in the Journal of Environmental Psychology

Word count: 7,404 (w/o Appendices)

# How do we know we are measuring environmental attitude?

## Specific objectivity as the formal validation criterion for measures of latent attributes

Florian G. Kaiser[1], Martin Merten[1], and Eunike Wetzel[2]

[1]Otto-von-Guericke University, Magdeburg, Germany

[2]University of Konstanz, Konstanz, Germany

**Author Note**

This research was financially supported by grant 03SFK4Q0 from the German Federal Ministry of Education and Research (BMBF) as part of the Copernicus Project ENavi. Collection of the data for this research was financially supported by grant TH-15/99-1 from the Swiss Federal Institute of Technology in Zürich, Switzerland. This article portrays exclusively the viewpoints of the authors.

We wish to thank Jane Zagorski for her language support, Livia Kaiser for her support in combinatorics, and Oliver Taube, Karen Draney, Gary Evans, Antal Haans, Terry Hartig, Laura Henn, Liane Hentschke, Alexandra Kibbe, Michael Knuth, Siegmar Otto, Mark Wilson, and three anonymous reviewers for their comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Florian G. Kaiser, Otto-von-Guericke University, Institute of Psychology, P.O. Box 4120, D-39016 Magdeburg, Germany; e-mail: florian.kaiser@ovgu.de

## Abstract

In this article, our aim is to introduce the formal validation criterion of *specific objectivity* to environmental psychologists. Using a convenience sample of 787 respondents, we constructed a set of attitude measures. We found that all five scales represented one attribute. As expected from the *Campbell Paradigm*, individual attitude can be *inferred* from verbal acts, such as expressions of appreciation, normative expectations, intentions to engage, and self-reports of past engagement in environmentally protective behaviors. Our findings challenge (a) the common practice in psychology of defining measures by their indicators and (b) the presumption of discriminant validity between typical measures of attitudes, subjective norms, perceived behavioral control, and intentions in research within the planned behavior framework. According to our research, the same validation criterion—specific objectivity—that is applied to the measurement of latent attributes in other empirical sciences (e.g., physics) can be applied to the measurement of attitudes.


*Keywords*: attitude measurement, planned behavior, test validity, environmental attitude, conservation (ecological behavior)

## 1. Introduction

An empirical science can cumulatively learn from observations only if it overcomes the specifics of its observations by revealing that there is an *objectively recognizable reality* behind these observations (Rasch, 1977). This is particularly challenging with latent attributes, which are only indirectly observable (e.g., people's environmental attitude or the temperature of objects). In environmental psychology, similar to psychology as a whole, many seem to believe in Stevens' (e.g., 1946) operational, rule-based definition, which states that measurement is the assignment of numbers to objects according to some rule. In such a rule-based measurement approach—also called measurement by fiat—measures depend on *theoretically postulated* relations between observable indicators and latent attributes (see Whitely, 1977).[1] With such an approach, however, the measurement of latent attributes is caught in what De Houwer, Gawronski, and Barnes-Holmes (2013) call a catch-22: The valid measurement of a latent attribute (e.g., attitude) requires knowledge about its valid indicators, and vice versa, identifying these indicators already necessitates the valid measurement of said attribute.

In our article, we argue that the particulars of any measurement instrument—and, thus, of its specific observations—can be overcome with *specific objectivity* (e.g., Rasch, 1977) or *scale-free* (e.g., Michell, 1986) and *sample-free* measurement (e.g., Fischer, 1987). Specific objectivity in psychology means that two people can be quantitatively compared with each other with regard to a latent attribute (e.g., environmental attitude) even if different measurement instruments have been used to assess the attribute. By demonstrating the *invariance* of this person comparison from the specifics of a measurement instrument, the reality of the underlying latent attribute becomes objectively recognizable (see e.g., Rasch, 1961; Wright, 1968).

In contrast to other empirical scientists (e.g., physicists), psychologists have to a large degree failed to embrace specific objectivity (see e.g., Fischer, 1987; Rasch, 1977) as the formal validation criterion for their concept measures. Thus, one price we probably pay as an empirical science is that, far too often, we seem to end up with irreproducible, false-positive findings (see e.g., Open Science Collaboration, 2015; Simmons, Nelson, & Simonsohn, 2011). Another price we probably pay—for concept measures that can be distinguished by name rather than by an objectively recognizable reality, by their “essence” so to speak—is that behavior explanation (e.g., as planned behavior; e.g., Ajzen, 1991) ends up representing more fiction than fact.

With this article, we aim to introduce specific objectivity in the measurement of environmental attitude to environmental psychologists, without the ambition of covering attitude or environmental attitude conceptions in their entirety. Specifically, we will illustrate how the existence of environmental attitude can be formally validated. From a methodological point of view, what we are presenting here is not a novelty (see e.g., Fischer, 1987; Rasch, 1977; Wright & Panchapakesan, 1969).

We will begin our illustration with the prototype of a latent attribute in an empirical science: temperature and its measurement. Subsequently, we will apply the same logic to attitude measurement. The environmental attitude measures that we use in our research are grounded in a paradigm for attitude research, which Kaiser and colleagues call the Campbell Paradigm (e.g., Kaiser, Byrka, & Hartig, 2010; Kaiser, Hartig, Brügger, & Duvier, 2013). We chose to focus on the Campbell Paradigm because it is one of the few (if not the only) attitude conception in social psychology that specifically requires the employment of a Rasch-family model (i.e., a one-parameter logistic item response theory model; see, e.g., Embretson & Reise, 2000) in the measurement of attitude. Before we turn to our specific hypotheses, we will briefly explain the relevant aspects of the Campbell Paradigm. Our research will allow readers to recognize the critical challenges in future environmental attitude research that still await scholars’ continuous efforts. We begin by providing an example of what specific objectivity means in connection with measurement.

### 1.1. Formal validation of *temperature* measures

The act of measurement in an empirical science begins with the discovery of units of measurement (e.g., a degree in Celsius [°C] or Fahrenheit [°F]; e.g., Michell, 1999). Subsequently, objects are assessed with respect to how many such units of the attribute they possess. Table 1 provides examples of five objects that incorporate different magnitudes (i.e., numbers of units) of temperature.

Table 1

*Temperature measures of some selected objects—measured in degrees Celsius (*°C*) and in degrees Fahrenheit (*°F*)—and some selected difference ratios.*

| °C | | | °F | | | Ratios of differences | | | |
|---|---|---|---|---|---|---|---|---|---|
| x | y | \|x-y\| | x | y | \|x-y\| | \|x-y\|:\|x-y\| | *in °C* | *in °F* | % |
| (a) 37.1 | (e) 5.8 | #1: 31.3 | (a) 97.9 | (e) 42.1 | #1: 55.8 | #1:#2 | 2.16 | 2.11 | 2.1 |
| (b) 22.6 | (a) 37.1 | #2: ***14.5*** | (b) 71.5 | (a) 97.9 | #2: ***26.4*** | ***#2:#4*** | **1.84** | **1.90** | 3.4 |
| (c) -4.1 | (b) 22.6 | #3: 26.7 | (c) 24.8 | (b) 71.5 | #3: 46.7 | #3:#1 | 0.85 | 0.84 | 1.9 |
| (d) 3.8 | (c) -4.1 | #4: ***7.9*** | (d) 38.7 | (c) 24.8 | #4: ***13.9*** | #4:#5 | 3.95 | 4.09 | 3.4 |
| (e) 5.8 | (d) 3.8 | #5: 2.0 | (e) 42.1 | (d) 38.7 | #5: 3.4 | #5:#2 | 0.14 | 0.13 | 6.9 |

*Note*. x and y represent the temperature of (a) a person, (b) a living room, (c) a beaker with ice cubes, (d) a veranda at night, and (e) a veranda at noon. |x–y| stands for the absolute difference between two measures. |x-y|:|x-y| gives a ratio of two differences. For example, #2:#4 refers to Difference #2 divided by Difference #4 for either the Celsius (14.5°C and 7.9°C) or the Fahrenheit (26.4°F and 13.9°F) measure. The resulting ratios of differences for Celsius (Fahrenheit) are depicted in the column labeled *in °C* (*in °F*). The last column (%) stands for the relative difference between the two ratios (relative to the average of the two ratios). All Fahrenheit (°F) measures were taken with the digital thermometer from TFA-Dostmann Kat. Nr. 30.1012. All Celsius (°C) measures were taken with the digital thermometer from TFA-Dostmann Kat. Nr. 30.1027. Both numbers refer to official catalogue numbers of products manufactured by TFA-Dostmann GmbH & Co. KG (tfa-dostmann.de).

In measurement, numbers are supposed to represent quantitative realities (e.g., Michell, 1999; Rasch, 1977). With interval-scale measures (e.g., the Celsius scale), these realities are *ratios of differences in magnitudes* of the attribute incorporated in objects. Our prototypical example in Table 1 shows five arbitrarily selected ratios of differences between pairs of measures of the temperatures of five objects. For example, the difference between the temperature of a warm living room (22.6°C/71.5°F) and a person's body temperature (37.1°C/97.9°F) is *14.5* in °C or *26.4* in °F; another difference is the one between the outside temperature on a February night (3.8°C/38.7°F) and the temperature inside a beaker filled with ice cubes (-4.1°C/24.8°F), which is *7.9* in °C or *13.9* in °F (see Table 1). From Table 1, we can further conclude that the ratio of these two differences (1.84 in °C/1.90 in °F) remains invariant irrespective of which temperature measure (Celsius or Fahrenheit) we use (see the columns labeled *in °C* and *in °F* in Table 1). In other words, the ratio of differences is (apart from some inaccuracies when reading the scales) a *constant* and, thus, an *objectively recognizable reality*.

According to Michell (1986; see also Rasch, 1977), the act of measurement is supposed to uncover such objectively recognizable numerical realities. Depending on the level of measurement, this can be either ratios of magnitudes (ratio scales) or ratios of differences in magnitudes (interval scales) of attributes incorporated in people or objects. These numerical facts are revealed as objectively recognizable realities when they remain invariant, even when we use completely distinct measurement instruments (e.g., Celsius and Fahrenheit scales) that are intended to measure the same attribute (i.e., temperature). In turn, the equivalence of the two instruments can be reduced to two unique linear regression equations that convert one measure of temperature into the other:

$$°C = 5/9\,(°F - 32) \tag{1}$$

$$°F = 9/5\,°C + 32 \tag{2}$$

In other words, irrespective of the measurement instrument (e.g., the Celsius or Fahrenheit scale), numerical relations represent the formal validation criterion by which one can corroborate the existence of the latent attribute of interest (e.g., temperature) and simultaneously validate instruments as measures of the latent attribute. However, by itself, independence from the measurement instrument does not validate the chosen label or conceptual understanding of the attribute. We will return to this issue in the Discussion.

Not surprisingly, long before temperature was theoretically understood in physics and thus recognized as a misnomer, temperature could be validly represented in distinctive interval-scale measures, each employing a unique unit (i.e., a degree in Fahrenheit, Celsius, Delisle, or Réaumur; see Chang, 2007). The first lesson to be learned from this historical excursion into physics is that, in psychology, we should be less concerned about specific indicators, attribute labels, and concept definitions than about the numerical relations of magnitudes of psychological attributes incorporated in people. These relations—and not the specific indicators—should remain *invariant* irrespective of the measurement instrument that is employed.

## 1.2. Measurement of environmental attitude

The most common notion of (explicit) attitudes is that attitudes represent latent attributes that are *manifested* when a person professes larger or smaller amounts of appreciation for an attitude object (e.g., Eagly & Chaiken, 1993; Rosenberg & Hovland, 1960). In this tradition, an attitude becomes obvious when people express larger or smaller amounts of esteem for an attitude object (e.g., environmental protection) or an attitude-implied behavioral goal (e.g., protecting the environment). Typically, these expressions can take three generic forms: (a) cognitive—when a person verbally admits to being an environmentalist or approves of pro-environmental behaviors or policies; (b) affective—when a person's face darkens in sadness upon hearing about the victims of an oil spill; (c) behavioral—when a person engages in pro-environmental activities or reports past pro-environmental activities. By far the most typical way to measure (explicit) attitudes is by means of cognitive and affective responses in the form of evaluative and normative statements in questionnaires (e.g., Krosnick, Judd, & Wittenbrink, 2005). Examples from environmental psychology can be found, among others, in Milfont and Duckitt (2004), Schultz (2001), and Thompson and Barton (1994).

In this measurement tradition, the item type typically reveals the concept or attribute to be measured. For example, behavior items, such as self-reports of past environmentally protective behavior and overt expressions of such behavior, are the logical foundation of pro-environmental behavior measures (see e.g., Kormos & Gifford, 2014) but not of people's environmental attitude. Inevitably, attitude measurement is caught in an intractable problem: The valid measurement of an attitude requires knowledge about its proper indicators, and vice versa, identifying these indicators already necessitates the valid measurement of said attitude (see De Houwer et al., 2013).

### 1.3. The Campbell Paradigm

Kaiser et al. (2010) proposed an alternative idea in attitude measurement. Their notion was named the Campbell Paradigm after Donald T. Campbell (1963), who had proposed the original conceptual idea. He argued that individual attitudes can and should be measured by indicators that are ordered with respect to their difficulty. For example, an environmentalist (i.e., a person who aspires to protect the environment and who, one might assume, holds a pronounced pro-environmental attitude) is typically expected to engage in a set of activities that reflect his or her attitude. For instance, she or he may publicly acknowledge that climate change is caused by humans, vote for representatives with a known pro-environmental record, recycle cardboard regularly, and eliminate foods that are particularly environmentally harmful (e.g., meat) from her or his diet. Generally, the person's esteem for an attitude object (e.g., environmental protection) or for an attitude-implied behavioral goal (e.g., protecting the environment) becomes obvious in the extent to which she or he engages in more and more difficult behaviors that involve increasingly demanding barriers or progressively more taxing sacrifices (i.e., increasing behavioral costs).

Consistent with Campbell’s conceptualization, Kaiser et al. (2010) maintain that the order of the difficulty of items is paramount in attitude measurement. Accordingly, and relevant to this article, they proposed that the Rasch family of models[2] (see e.g., Rasch, 1960/1980) be used to measure attitudes. For the Campbell Paradigm, it is essential to order the indicators of an attitude by their difficulty level. However, it is not essential for the indicators from which an attitude is inferred to be exclusively comprised of cognitive or affective responses in the form of evaluative and normative statements. Behavioral responses in the form of reports of past behavior or overt acts work as well. Expectedly, Byrka and Kaiser (2013) confirmed that traditional attitude items in the form of verbal expressions of appreciation for behaviors, along with self-reports of behavior, can be combined into a single category of indicators (see also Brügger et al., 2011).

So far, the Campbell Paradigm has inspired various Rasch-model-based attitude measures: attitude toward environmental protection (i.e., environmental attitude; see Kaiser et al., 2013), attitude toward nature (see Brügger et al., 2011), attitude toward global climate change (see Urban, 2016), health attitude (see Byrka & Kaiser, 2013), and acceptance of (or attitude toward) some specific restrictions (see Byrka, Kaiser, & Olko, 2017).

### 1.4. Research goals

Our research is aimed at illustrating that specific objectivity is the ultimate formal validation criterion in an empirical science. We thus aim to corroborate the idea that the type of indicator is not a defining feature of environmental attitude and can therefore not be used to validate environmental attitude measures. On the contrary, along with evaluative and normative statements that address the same attitude, self-reports of past behavior are indicators that can be used to measure people's attitudes as well.

First, we identified 50 self-reports reflecting past environmentally protective behaviors and 48 evaluative and normative statements concerning the same environmentally protective behaviors (i.e., attitude toward behaviors, subjective norms, perceived behavioral control, and behavioral intentions) that are commonly used in the assessment of the planned behavior components (see e.g., Ajzen & Madden, 1986). Second, we used these 98 statements to develop five Rasch scales that are supposed to measure environmental attitude (i.e., attitude toward environmental protection). We expected these five measures to reflect one latent attribute when we explored their convergent and discriminant validity.

Of our five attitude measures, two represent combinations of 24 evaluative and normative statements and 25 behavioral self-report items without a single overlapping item. These two perfectly distinct and, therefore, independent instruments were subsequently used to establish a unique linear regression line that can be used to convert one environmental attitude measure into the other. Along this regression line, we can locate people who incorporate magnitudes of the latent attribute that we believe represents environmental attitude. These magnitudes in turn can be used to expose invariant, and, thus, objectively recognizable numerical realities that are independent of each specific measurement instrument: the formal validation criterion of our measures of *environmental attitude*. Note that the very label (e.g., environmental attitude) refers to the theoretically presumed attribute of interest; whether or not the chosen label is accurate has to be confirmed empirically.

## 2. Method

### 2.1. Participants

Our sample was randomly selected in 2001 from the resident registers of four communities in Saxony-Anhalt, one of Germany's 16 states. The communities differed in the absolute numbers of inhabitants. Of the 5,000 people who were selected, 787 (58% women, *n* = 453; 42% men, *n* = 328) returned completed questionnaires (response rate: 16%). The participants' median age was 46 (*M* = 46.2, *SD* = 16.4; range: 18 to 89). Due to the relatively high nonresponse rate, our participants could not be considered representative of the adult population of Saxony-Anhalt, but representativeness was not crucial here. It was sufficient that there were no severe range restrictions in our measures in this sample (see Table AI-1 in Appendix I).

### 2.2. Measurement instruments

The data for our research were originally collected for a study in the planned behavior framework (see Kaiser, Schultz, & Scheuthle, 2007: Study 2). The questionnaire consisted of 50 behavioral self-report items and four sets of 12 evaluative statements, with each set representing the conventional planned behavior variables: attitude toward behavior, subjective norms, perceived behavioral control, and behavioral intention. Each of these 48 statements corresponded to one of the behavioral self-report items. The two behaviors that were left over were *I am a member of a carpool* and *I requested an estimate for the installation of solar power* (for all items and their original usage, see Kaiser et al., 2007).

By calibrating five environmental attitude (EA) measures using behavioral self-reports, evaluative statements, or both, we tested our basic idea that all these items represent one underlying latent attribute. One measure was based exclusively on the 50 behavioral self-reports (EA-50). Another was based on the 48 evaluative statements (EA-48). A third measure was a combination of the 50 self-reports and 48 evaluative statements (EA-98), and the fourth and the fifth were two evenly split versions of 49 items each (EA-49A and EA-49B) from the EA-98 scale. Each of the two EA-49 measures consisted of 25 behavioral self-reports and 24 evaluative statements. The pairs of split measures (i.e., EA-50 and EA-48; EA-49A and EA-49B) did not contain any overlapping items. As all five measures were calibrated as Rasch scales, they should, in principle, represent specifically objective attitude measures (see e.g., Fischer, 1987; Rasch, 1977; Wright & Panchapakesan, 1969). For example items and further information on instrument calibrations, see Appendix I.

## 3. Results

We report our findings in three sections. In Section 1, we review the construct validity and the psychometric features of five specific-objectivity-based attitude measures. Moreover, we establish the linear relation between two of the completely distinct measurement instruments that were both intended to measure the latent attribute environmental attitude (i.e., EA-49A and EA-49B). In Section 2, we report the evidence for the objectively recognizable reality of *environmental attitude* that goes beyond any particular instrument designed for its measurement. In Section 3, we contrast the findings from Section 2 with findings from instruments that are constructed in a rule-based tradition with predetermined metrics or units, respectively (i.e., z-scores).

### 3.1. Five measures, five metrics — One attribute

The five measures of environmental attitude were all strongly correlated (Pearson's $r$ at least $r = .65$), even when they were not corrected for measurement error attenuation (see Table 2). When corrected for attenuation, all but two of the 10 correlations were perfect.[3]

The two *corrected* correlations that were not perfect (i.e., $r < 1.0$) involved completely distinct measures with no overlapping indicators. Specifically, the correlation between the measure that was comprised of behavioral self-reports (EA-50) and the measure that was comprised of evaluative statements (EA-48) was $r = .83$; and the correlation between the two distinct EA-49 measures was $r = .94$.

Table 2

*Descriptive statistics, reliabilities, and bivariate correlations of five different estimates of people's environmental attitude.*

| | *N* | *M* | *SD* | EA-50 | EA-48 | EA-98 | EA-49A | EA-49B |
|---|---|---|---|---|---|---|---|---|
| Self-report items (EA-50) | 787 | -.09 | .81 | *.74* | .83 | 1.00[a] | 1.00[a] | 1.00[a] |
| Evaluative statements (EA-48) | 783 | .39 | .89 | .65 | *.83* | 1.00[a] | 1.00[a] | 1.00[a] |
| Combined (EA-98) | 787 | .14 | .77 | .87 | .94 | *.88* | 1.00[a] | 1.00[a] |
| Version A (EA-49A) | 787 | .12 | .79 | .82 | .85 | .93 | *.76* | .94 |
| Version B (EA-49B) | 787 | .16 | .87 | .80 | .89 | .94 | .74 | *.81* |

*Note*. *M*s and *SD*s are expressed in logits, the conventional units of Rasch scales. In the five presented instances, they ranged from -3 to +4 logits. Rasch-based measures are typically interval scales. In the correlation matrix (columns EA-50 to EA-49B), the diagonal cells indicate separation reliability estimates (for details, see Wright & Masters, 1982). Off-diagonal values represent Pearson correlations that are either uncorrected for measurement error attenuation (below the diagonal; all coefficients are statistically significant at $p < .001$) or corrected (above the diagonal). [a]Corrections resulting in correlations exceeding 1.00 were truncated to 1.00.

Table 2 furthermore reveals a clear lack of evidence for discriminant validity as even the *uncorrected* correlations between the five measures of environmental attitude matched or exceeded the five measures' reliabilities (see Campbell & Fiske, 1959). In other words, all five measures seemed to reflect the same psychological attribute irrespective of the differences in the item sets that were used.

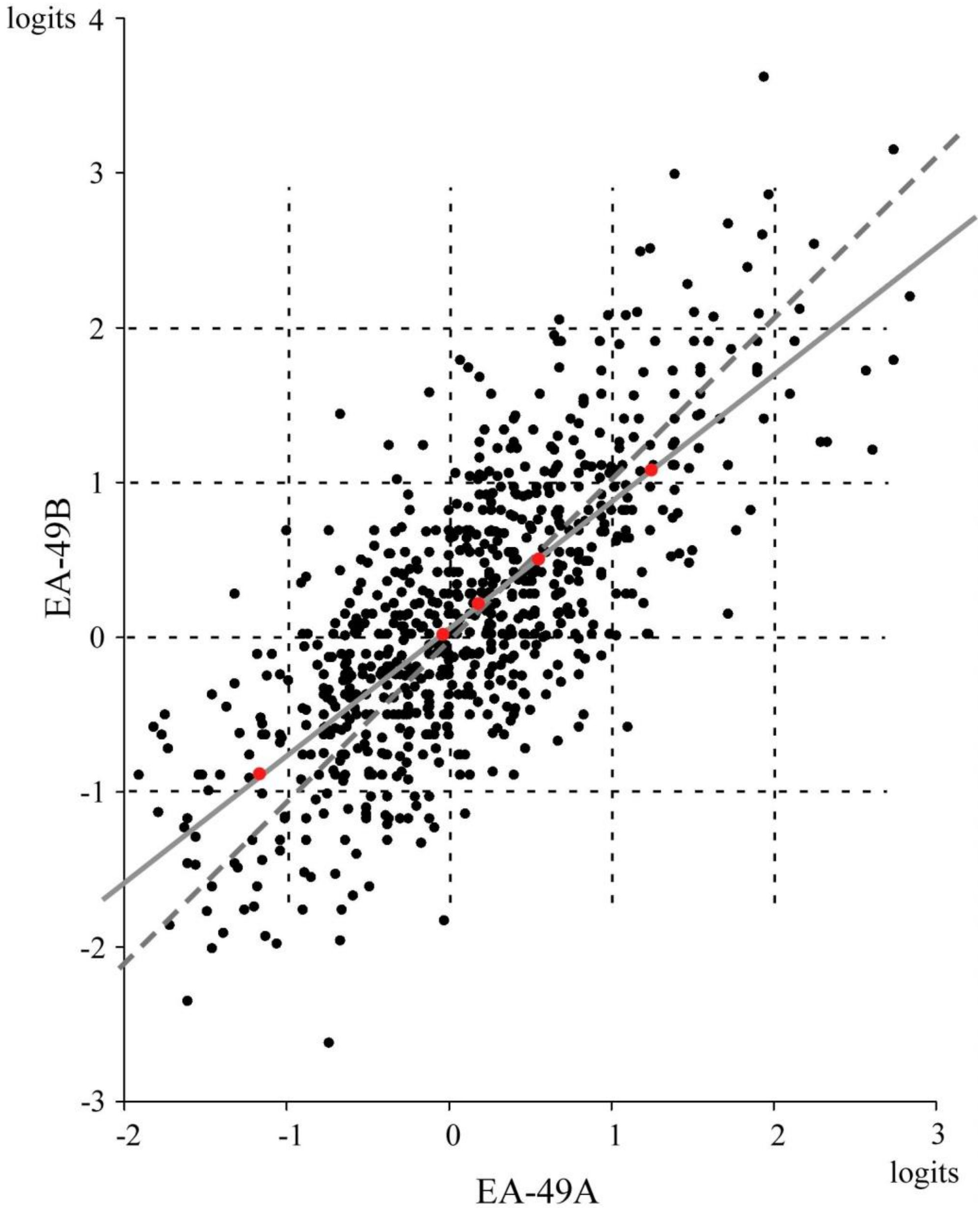


*Figure 1.* Scatterplot of individual environmental attitude estimates represented in logits for two specific-objectivity-based measures (EA-49A and EA-49B). The solid regression line is based on logits, whereas the dashed regression line (approximated for illustrative purposes) is based on z-scores. *N* = 787. Red dots represent the coordinates of the five people (a through e) selected for Table 3—Panel A.

The dissimilar means and standard deviations in Table 2 reflect the distinct metrics of the various measurement instruments. This distinction could also be seen when EA-49B was regressed on EA-49A: EA-49B = .82*EA-49A + 0.06 (the solid regression line in Figure 1). Even though the units are called logits for both specific-objectivity-based instruments, the actual size of a unit differed between the two measures. One logit in EA-49B was somewhat larger than one logit in EA-49A. Whereas unit differences are not a problem for interval scales (see e.g., degrees in Fahrenheit vs. degrees in Celsius), by contrast, using the same label for different unit sizes can create confusion. Next, we show how two of our *environmental attitude* measures were able to reflect the objectively recognizable reality of invariant difference ratios of people's latent attitude levels.

### 3.2. Specific-objectivity-based measures of *environmental attitude*

The objectively recognizable reality of an attribute assessed with two instruments can be reduced to a unique linear regression equation that converts—as with Fahrenheit and Celsius measures of temperature (see Formulas 1 & 2)—one measure of *environmental attitude* into the other. Accordingly, only data points *on* the regression lines in Figure 1 can represent people for whom two measures of environmental attitude are specifically objective. In other words, only values on the regression lines can expose objectively recognizable numerical realities: i.e., invariant ratios of differences.

With about one fourth to one fifth of the variance representing measurement error (i.e., 24% [EA-49A] and 19% [EA-49B]) in our person estimates, measurement is far from what is standard for measurement accuracy in other empirical sciences (e.g., the measurement of temperature in physics). This comparatively unreliable measurement of attitude is reflected by the fuzzy scatterplot of the two person estimates (see Figure 1; for numerical reliabilities, see Table 2).

To demonstrate that the specifically objective measurement of people is possible in principle, we selected five people whose scores fell on the solid regression line in Figure 1 and were spread along the entire continuum. The coordinates of the five people with respect to the two instruments, EA-49A and EA-49B, were: -1.17/-0.89, -0.04/0.02, 0.18/0.21, 0.60/0.55, 1.24/1.08 (see Table 3—Panel A).

Table 3

*Specific-objectivity-based (Panel A) or rule-based (Panel B) person estimates of environmental attitude—measured with EA-49A and EA-49B—and five selected difference ratios.*

A

| EA-49A | | | EA-49B | | | Ratios of differences | | | |
|---|---|---|---|---|---|---|---|---|---|
| x | y | \|x-y\| | x | y | \|x-y\| | \|x-y\|:\|x-y\| | *EA-49A in logits* | *EA-49B in logits* | % |
| (a)-1.17 | (e) 1.24 | #1: 2.41 | (a)-0.89 | (e) 1.08 | #1: 1.97 | #1:#2 | 2.13 | 2.16 | 1.5 |
| (b)-0.04 | (a)-1.17 | #2: ***1.13*** | (b) 0.02 | (a)-0.89 | #2: ***0.91*** | ***#2:#4*** | **2.69** | **2.68** | 0.5 |
| (c) 0.18 | (b)-0.04 | #3: 0.22 | (c) 0.21 | (b) 0.02 | #3: 0.19 | #3:#1 | 0.09 | 0.10 | 5.5 |
| (d) 0.60 | (c) 0.18 | #4: ***0.42*** | (d) 0.55 | (c) 0.21 | #4: ***0.34*** | #4:#5 | 0.66 | 0.64 | 2.3 |
| (e) 1.24 | (d) 0.60 | #5: 0.64 | (e) 1.08 | (d) 0.55 | #5: 0.53 | #5:#2 | 0.57 | 0.58 | 2.8 |

B

| EA-49A | | | EA-49B | | | Ratios of differences | | | |
|---|---|---|---|---|---|---|---|---|---|
| x | y | \|x-y\| | x | y | \|x-y\| | \|x-y\|:\|x-y\| | *EA-49A in z-scores* | *EA49B-in z-scores* | % |
| (a)-1.30 | (e) 1.12 | #1: 2.42 | (a)-0.96 | (e) 0.82 | #1: 1.78 | #1:#2 | 2.80 | 2.75 | 1.8 |
| (b)-0.43 | (a)-1.30 | #2: ***0.86*** | (b)-0.31 | (a)-0.96 | #2: ***0.65*** | ***#2:#4*** | **5.00** | **4.00** | 22.2 |
| (c) 0.26 | (b)-0.43 | #3: 0.69 | (c) 0.18 | (b)-0.31 | #3: 0.49 | #3:#1 | 0.29 | 0.27 | 4.7 |
| (d) 0.43 | (c) 0.26 | #4: ***0.17*** | (d) 0.34 | (c) 0.18 | #4: ***0.16*** | #4:#5 | 0.25 | 0.33 | 28.6 |
| (e) 1.12 | (d) 0.43 | #5: 0.69 | (e) 0.82 | (d) 0.34 | #5: 0.49 | #5:#2 | 0.80 | 0.75 | 6.5 |

*Note*. Each of the five person estimates represents a single person (a through e). |x–y| stands for the absolute difference between two person estimates. |x-y|:|x-y| specifies the two differences in a ratio. In Panel A, for example, #2:#4 refers to Difference #2 divided by Difference #4 either in logits of the EA-49A measure (1.13 and 0.42) or in logits of the EA-49B measure (0.91 and 0.34). The resulting ratios of differences are in the columns labeled *EA-49A in logits* or *EA-49B in logits.* In Panel B, #2:#4 refers to Difference #2 divided by Difference #4 for both the EA-49A measure (0.86 and 0.17) and the EA-49B measure (0.65 and 0.16) both expressed as z-scores. The resulting ratios of differences are in the columns labeled *EA-49A in z-scores or EA-49B in z-scores.* The last column (%) stands for the relative difference between the two ratios (relative to the average of the two ratios).

The creation of five combinations of two differences out of 10 possible differences results in 45 unique ratios. These 45 difference ratios between EA-49A and EA-49B differed on average by 2.1% (see Table AII-1 in Appendix II). Only six (13.3%) of the 45 difference ratios established with the two instruments exceeded 5.0%, with a maximum difference of 7.0%. Such fairly invariant difference ratios point to the measurement-instrument-independent existence of a latent attribute that these five people possess (among other things). As such, these ratios are evidence of an objectively recognizable reality that transcends the specific measures (i.e., EA-49A or EA-49B; see Table 3—Panel A).

3.3. Rule-based measures of *environmental attitude*

We would have come to quite a different conclusion if we had created rule-based rather than specific-objectivity-based measures of *environmental attitude* by imposing units or metrics rather than by empirically searching for them. To demonstrate this, we formed rule-based measures by z-transforming the sum scores of EA-49A and EA-49B. We then regressed EA-49B on EA-49A again. This regression line, based on such rule-based units, was different from the one based on logits: EA-49B = .74*EA-49A (the dashed regression line in Figure 1). This time, with imposed uniform metrics (z-scores), the units for EA-49A and EA-49B were equal in size, and the regression line had no intercept (the dashed regression line in Figure 1).

As before, we selected five people whose scores fell on the regression line—the dashed regression line in Figure 1—and were spread along the entire continuum. The coordinates of the five selected individuals were: -1.30/-0.96, -0.43/-0.31, 0.26/0.18, 0.43/0.34, and 1.12/0.82 (see Table 3—Panel B). Their 45 difference ratios differed on average by 7.3%. Of the 45 difference ratios, 17 (37.8%) even exceeded 5.0% with a maximum difference of 28.6% (see Table AII-1 in Appendix II). This time, the difference ratios appeared less invariant (see Table 3—Panel B). What we call *environmental attitude* here does not seem to transcend the specific measurement instruments. The two rule-based *environmental attitude* measures seem to indicate two instrument-specific attitudes underlying people’s responses. With an even more rule-based measure (e.g., using a 1 to 10 scoring of people within each of the 10 sum score percentiles instead of z-transforming people's sum scores), the level of divergence across the 45 difference ratios would have been even more pronounced.

## 4. Discussion

In measurement, numbers are supposed to represent objectively recognizable realities (e.g., Michell, 1999; Rasch, 1977). In psychology, these realities are invariant (i.e., independent of the measurement instrument) difference ratios of magnitudes of latent attributes that people possess. In turn, the existence of such invariant ratios formally validates the existence of the latent attribute (e.g., *environmental attitude*). Our demonstration in this article indicates that environmental attitude possesses such a reality that can be made objectively recognizable with two distinct but substantially correlated attitude measures. Such attribute measures do not necessarily need a theoretical or practical rationale, not even face validity. However, this is not to say that the mere existence of invariant difference ratios of magnitudes of an attribute already disclose what the latent attribute represents: its essence.[4]

In and of itself, validating a reality behind the numbers assigned to people is not sufficient to corroborate the essence of the attribute an instrument is presumed to measure (see e.g., Whitely & Dawis, 1974). To gain a thorough understanding of an attribute, we must apply external criteria from *outside the measurement process* (Whitely, 1977), ideally in experimental research.[5] For the current example, the proposed *environmental attitude* measure was previously found to significantly predict people's energy consumption (see Arnold, Kibbe, Hartig, & Kaiser, 2017) and pro-environmental behavior (see Kaiser & Byrka, 2015), including a person's membership in an environmental organization (see Arnold & Kaiser, 2016). Thus, we can be fairly confident that we are measuring something akin to a person's appreciation for environmental protection (i.e., the person's environmental attitude).

With our data, which were originally collected for a study in the planned behavior framework (see Kaiser et al., 2007), we found bivariate correlations that exceeded the reliability estimates of the corresponding environmental attitude measures (see Table 2). Such a finding provides evidence for a lack of discriminant validity among these measures (see Campbell & Fiske, 1959). Even the measures with no overlapping items (i.e., the measure based on self-reports of behavior [EA-50] and the measure based on evaluative statements [EA-48], as well as the two distinct EA-49 measures consisting of 25 self-reports and 24 evaluative statements) were strongly correlated ($r$ = .83 and $r$ = .94, respectively), particularly when corrected for measurement error attenuation. We can conclude from these results that the type of indicator—evaluative statement or behavioral self-report—is not a defining feature of an attitude measure, a finding that is in line with the classical notion of attitude measurement, the so-called tripartite model (see e.g., Eagly & Chaiken, 1993; Rosenberg & Hovland, 1960).

As people's attitudes can become objectively recognizable only in people's behavior (De Houwer et al., 2013), people's attitudes must be inferred from appraisals and reports of past behaviors (i.e., verbal behavior) and from actual behaviors observed in labs or in real-life situations (see Kaiser et al., 2010). Accordingly, four of our five attitude measures were comprised of indicators that are typically used to measure attitudes toward behaviors, subjective norms, perceived behavioral control, and behavioral intentions (cf. e.g., Ajzen & Madden, 1986).

On the one hand, our findings make common practice in research within the planned behavior framework look problematic because the typical measures of attitudes, subjective norms, perceived behavioral control, and intentions obviously lack discriminant validity as their items could be combined into a single Rasch scale (see Tables 2 & AI-1 in Appendix I). As such, our findings confirm the large correlations often found between planned behavior concepts that are believed to be discriminant measures given their distinct labels: attitude toward behavior, subjective norms, perceived behavior control, intention, and behavior (see e.g., Kaiser et al., 2007).

On the other hand, our findings also provide support for the Campbell Paradigm (see Kaiser et al., 2010), namely, for the notion that personal attitudes can be *inferred* from verbal acts, such as expressions of appreciation for, perceived normative and control expectations regarding, intentions to engage in, and self-reports of past engagement in environmentally protective behaviors. In summary, self-reports of past behavior and evaluative statements addressing one and the same attitudinal object (e.g., environmental protection) are all verbal indicators of people's environmental attitude (see e.g., Kaiser & Byrka, 2015).

Moreover, our findings call into question common practices in psychological research where the demands for innovation and complete understanding tend to result in a continual proliferation of novel attributes and instruments. Out of this necessity, psychological researchers often rely on proxies or on measures with only a few indicator items (typically even fewer than seven; see e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999). Yet, when even two attitude measures, each consisting of nearly *50 indicators*, share only about half of their variance (i.e., 54.8%; derived from $r$ = .74 in Table 2), we obviously face a dilemma (i.e., choosing between the comprehensiveness of behavior explanation and measurement accuracy) that seems rather intractable.

Of course, such a limited level of accuracy in measures does not allow for the valid distinction of individuals who possess very similar quantities or magnitudes of a latent attribute. Still, such moderate accuracy figures indicate that there are at least two instruments that are appropriate enough for the empirical exploration of the theoretically anticipated connections between environmental attitude, energy consumption, and various environmentally protective activities (see e.g., Arnold et al., 2017; Byrka et al., 2017). In turn, such empirical research will help us to cumulatively refine the corresponding theoretical and conceptual notions concerning environmental attitude and the corresponding measurement instruments in a way that is comparable to other empirical sciences (e.g., physics).

In physics, only the understanding of temperature as encapsulated kinetic energy eventually led to the discovery of an absolute zero point and subsequently to the development of the Kelvin scale, a ratio scale measure of *temperature*, or more accurately, of kinetic energy (see Chang, 2007). In contrast to physicists, psychologists cannot be expected to discover instances of complete absence of attributes. Thus, without such an absolute zero point, psychologists can presumably expect to attain only interval-scale measures. Despite the encouraging start in the right direction, the scatterplot of the two environmental attitude measures (see Figure 1) leaves plenty of room for technical advances if the measurement of people's environmental attitude is going to compete with the accuracy of temperature measures (see Table 1).

Our findings furthermore call into question the standard notion of measurement in psychology as the assignment of numbers to people according to rules (cf. Stevens, 1946). When the assigned magnitudes primarily reflect rules and not objective realities of sorts, we should also not be surprised when psychological research reflects fiction rather than fact; fiction that often leads to nonreplicable findings (see e.g., Open Science Collaboration, 2015; Simmons et al., 2011). The replicability, and in the long run, the relevance of our findings in psychological science require valid attribute (i.e., concept) measures. In turn, the validity of our measures depends on the discovery of magnitudes that represent quantitative, objectively recognizable numerical facts incorporated in people (e.g., Michell, 1986, 1999). As we argue, only specifically objective measures of latent attributes can serve as the foundation that is needed for a cumulative empirical psychological science.

## References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50,* 179-211.

Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology, 22,* 453-474.

Arnold, O., Kibbe, A., Hartig, T., & Kaiser, F. G. (2017). Capturing the environmental impact of individual lifestyles: Evidence for the criterion validity of the General Ecological Behavior scale. *Environment and Behavior.* Advance online publication. doi:10.1177/0013916517701796

Arnold, O., & Kaiser, F. G. (2016). Understanding the foot-in-the-door effect as a pseudo effect from the perspective of the Campbell paradigm. *International Journal of Psychology.* Advance online publication. doi:10.1002/ijop.12289

Brügger, A., Kaiser, F. G., & Roczen, N. (2011). One for all? Connectedness to nature, inclusion of nature, environmental identity, and implicit association with nature. *European Psychologist, 16,* 324-333.

Byrka, K., & Kaiser, F. G. (2013). Health performance of individuals within the Campbell paradigm. *International Journal of Psychology, 48,* 986-999.

Byrka, K., Kaiser, F. G., & Olko, J. (2017). Understanding the acceptance of nature-preservation-related restrictions as the result of the compensatory effects of environmental attitude and behavioral costs. *Environment and Behavior, 49,* 487-508.

Campbell, D. T. (1963). Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 6, pp. 94-172). New York: McGraw-Hill.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Chang, H. (2007). *Inventing temperature: Measurement and scientific progress.* New York: Oxford University Press.

Charles, E. P. (2005). The correction for attenuation due to measurement error: Clarifying concepts and clarifying confidence sets. *Psychological Methods, 10,* 206-226.

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology, 24,* 252-287.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes.* Fort Worth, TX: Harcourt Brace Jovanovich.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272-299.

Fischer, G. (1987). Applying the principles of specific objectivity and of generalizability to the measurement of change. *Psychometrika, 52,* 565-587.

Kaiser, F. G., & Byrka, K. (2015). The Campbell paradigm as a conceptual alternative to the expectation of hypocrisy in contemporary attitude research. *The Journal of Social Psychology, 155,* 12-29.

Kaiser, F. G., Byrka, K., & Hartig, T. (2010). Reviving Campbell's paradigm for attitude research. *Personality and Social Psychology Review, 14,* 351-367.

Kaiser, F. G., Hartig, T., Brügger, A., & Duvier, C. (2013). Environmental protection and nature as distinct attitudinal objects: An application of the Campbell paradigm. *Environment and Behavior, 45,* 369-398.

Kaiser, F. G., Schultz, P. W., & Scheuthle, H. (2007). The theory of planned behavior without compatibility? Beyond method bias and past trivial associations. *Journal of Applied Social Psychology, 37,* 1522-1544.

Kormos, Ch., & Gifford, R. (2014). The validity of self-report measures of proenvironmental behavior: A meta-analytic review. *Journal of Environmental Psychology, 40,* 359-371.

Krosnick, J. A., Judd, C. M., & Wittenbrink, B. (2005). The measurement of attitudes. In D. Albaracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 21-76). Mahwah, NJ: Erlbaum.

Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin, 100,* 398-407.

Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept.* Cambridge, UK: Cambridge University Press.

Milfont, T. L., & Duckitt, J. (2004). The structure of environmental attitudes: A first- and second-order confirmatory factor analysis. *Journal of Environmental Psychology, 24,* 289-303.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349,* 943.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proceedings of the fourth Berkeley Symposium on mathematical statistics and theory of probability* (Vol. IV, pp. 321-333). Berkeley, CA: University of California Press.

Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. In M. Blegvad (ed.), *Danish yearbook of philosophy* (Vol. 14, pp. 58-93). Copenhagen, Denmark: Munksgaard.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* Chicago: University of Chicago Press (Original work published 1960).

Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In C. I. Hovland & M. J. Rosenberg (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (pp. 1-14). New Haven, CT: Yale University Press.

Schultz, P. W. (2001). The structure of environmental concern: Concern for self, other people, and the biosphere. *Journal of Environmental Psychology, 21,* 327-339.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359-1366.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103,* 667-680.

Thompson, S. C. G., & Barton, M. (1994). Ecocentric and anthropocentric attitudes toward the environment. *Journal of Environmental Psychology, 14,* 149-157.

Urban, J. (2016). Are we measuring concern about climate change correctly? Testing a novel measurement approach with the data from 28 countries. *Climatic Change, 139,* 397-411.

Whitely, S. E. (1977). Models, meanings and misunderstandings: Some issues in applying Rasch’s theory. *Journal of Educational Measurement, 14,* 227-235.

Whitely, S. E., & Dawis, R. V. (1974). The nature of the objectivity with the Rasch model. *Journal of Educational Measurement, 11,* 163-178.

Wright, B. (1968). Sample-free test calibration and person measurement. *Proceedings of the 1967 invitational conference on testing problems* (pp. 85-101). Princeton, NJ: Educational Testing Service.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement.* Chicago: MESA.

Wright, B., & Panchapakesan, N. (1969). A procedure for sample free item analysis. *Educational and Psychological Measurement, 29,* 23-48.

**Endnotes**

1 Rule-based measures are grounded in agreed upon established procedures or item sets within the so-called operational paradigm of measurement (see Michell, 1986).

2 As specific objectivity is typically a recognized feature of all one-parameter logistic item response theory models, our argument applies to the partial-credit Rasch model, the rating-scale Rasch model, the constrained graded-response model, and others alike. For an example that uses, for instance, the partial-credit Rasch model and thus an ordinal response format for the measurement of an attitude within the Campbell Paradigm, see Brügger, Kaiser, and Roczen (2011).

3 A generic correction for measurement error attenuation adjusts correlations by the reliabilities of the two measures involved (Charles, 2005). For this, one divides the uncorrected bivariate correlation by the square root of the product of the two reliabilities.

4 Without attribute measures reflecting an objective essence, even experimental research would at best result in social conventions (requiring mutually agreed upon interpretations of empirical findings) rather than objective laws describing actual and, thus, generalizable relations among (latent or manifest) attributes. Empirical science would be reduced to *a debate about facts* (i.e., the proper understanding of findings) rather than a *demonstration of facts* that can be repeated anytime at will—not even requiring *strict replications* of experiments under perfectly identical conditions.
For example, the relevant attributes—next to a constant (i.e., the gravitational constant)—involved in Newton’s law of universal gravitation are force, mass, and distance. The law describes the attractive force with which two masses (e.g., an apple and the Earth) attract each other. This force is proportional to the product of the two masses and inversely proportional to the squared distance between the two masses. Knowing the essence of this force allows the same explanation to be applied to rather diverse phenomena (e.g., gravity on earth and planetary movements). More important for our argument in this article, when our measures of mass, distance, and force validly reflect what distance, mass, and force are (i.e., reflect their essence), we are equipped to setup new experiments and forecast the force with which two other objects (e.g., a hypothetical elephant on planet Mars) attract each other. Essential to such *extended replications* is, however, the measurement of the three attributes, validly reflecting the essence of mass, distance, and force.

5 Accordingly, empirical scientists are supposed to explore the *causal* network in which a latent attribute is expected to operate. The notion of temperature as encapsulated kinetic energy, for example, logically leads to expectations about the thermic expansion of materials, material density, and electrical conductivity, the relations between which can be tested empirically and described mathematically.

## APPENDIX I

AI-1. Items

Typical examples of behavioral self-report items are *I reuse my shopping bags* and *I am a member of an environmental organization*. Engagement was verified with a yes/no format for 18 of the 50 behavioral self-report items and with a 5-point frequency scale (1 = *never,* 5 = *always*) for the remaining 32 self-report items. The responses to the latter set of items were recoded into a dichotomous format by collapsing *never*, *seldom*, and *occasionally* into *unreliable environmentally protective engagement*. *Often* and *always* were combined into *reliable environmentally protective engagement*. This step was taken because participants tend to use the response alternatives inconsistently, as our own research showed (see Kaiser & Wilson, 2000). Although unreliable, more nuanced response options are nevertheless preferred by respondents. Nineteen self-report items represented environmentally harmful activities; they were reverse coded. For all 50 items, *I cannot answer* was a response alternative that was coded as missing.

For the evaluative statements, attitude toward behavior was measured by asking participants to rate each of six behaviors on a 5-point response scale (1 = *bad*, 5 = *good*) and six other behaviors on a 5-point response scale (1 = *inappropriate*, 5 = *appropriate*). Subjective norms were measured by asking participants to rate each of six behaviors on a 5-point response scale (1 = *unlikely*, 5 = *likely*) with the opening phrase *Most people who are important to me think I should…* and six other behaviors on a 5-point response scale (1 = *disagree*, 5 = *agree*) with the opening phrase *Most people who are important to me….* Perceived behavioral control was also measured by asking participants to rate each of six behaviors on a 5-point response scale (1 = *difficult*, 5 = *easy*) and six other behaviors on a 5-point response scale (1 = *complicated*, 5 = *simple*). Intentions were measured by means of six behaviors on a 5-point response scale (1 = *unlikely*, 5 = *likely*) with the opening phrase *I* *will…* and six other behaviors on a 5-point response scale (1 = *undecided*, 5 = *decided*) with the opening phrase *I intend to…*. All responses were subsequently recoded into a dichotomous format by collapsing the two positive responses into a positive valuation of environmentally protective behavior and the two negative and the neutral response into absence of a positive valuation of environmentally protective behavior. Dichotomizing polytomous response scales is an established precaution for guarding against excessive measurement error in items requiring introspection and recollection, particularly in attitude research (for supporting evidence, see e.g., Matell & Jacoby 1971; Peabody, 1962) and is not the same as the controversial mean or median split technique occasionally used in experimental research (see DeCoster, Iselin, & Gallucci, 2009).

In the following section, we will summarize the calibration of five measures of *environmental attitude*, all of which are grounded in the Campbell Paradigm. Person scores were derived as maximum likelihood estimates expressed in logits, which are commonly used with Rasch-model-based instruments (for more details, see e.g., Embretson & Reise, 2000). Logits stand for the natural logarithm of the ratio between the probability of a positive and the probability of a nonpositive response across the entire response vector of a person. Logits represent units of an interval scale measure and can maximally range between $\pm\infty$ logits; our specific measures range from -3 to +4 logits. As with all interval scales, the zero point is arbitrary and is chosen for technical reasons. The smaller the logit value, the lower the particular person score. In our analyses, logits are derived by the Joint Maximum Likelihood (JML) estimation procedure. JML can in principle accommodate missing values (see e.g., Bond & Fox, 2007; Embretson & Reise, 2000).

Nevertheless, person scores could not be estimated when there was no variance in a data vector (e.g., all items rated with either a *positive* or *nonpositive* valence). The latter explains why we had four missing values in one of the five attitude measures (see Table 2). In the next section, we will summarize the calibration of the five measures of *environmental attitude*.

AI-2. Instrument calibrations

The results of all five Rasch model tests were in line with previous such calibrations (see e.g., Kaiser, Hartig, Brügger, & Duvier, 2013; Kaiser & Wilson, 2004). They by and large revealed reasonable and comparable statistical fit on the item and person levels. Because our sample of 787 participants was relatively large, we relied on the mean square (*MS*) statistic to assess fit. This fit measure reflects the relative discrepancy in variation between model prediction and observed data and is not strongly affected by sample size. For example, an average *MS* of 0.80 corresponds to a 20% lack of variation, and an *MS* value of 1.20 indicates 20% more variation in the data than what was predicted by the Rasch model. An *MS* value of 1.20 is commonly regarded as a sensible threshold for instruments used in the scientific exploration of empirical relations (cf. Wright, Linacre, Gustafson, & Martin-Löf, 1994).

The overall fit statistics for the items from the five scales were acceptable; *M(MS)* = 1.00 and *SD(MS)* between 0.05 and 0.06 (see Table AI-1). An acceptable item fit was to be expected because none of the items from any of the five models exceeded the threshold of fit that is deemed sensible (i.e., *MS* = 1.20). The overall fit statistics for people were acceptable as well; *M(MS)* = 1.00 and 1.01 and *SD(MS)* between 0.20 and 0.26. The acceptable person fit was also reflected by the comparatively small numbers of poorly fitting people ($t \geq 1.96$), which ranged from 36 (4.6%) to 73 (9.3%) across the five Rasch scales (see Table AI-1). Except for EA-98 (the all-inclusive scale) with 9.3%, the proportions of poorly fitting participants (6.6% or better) were otherwise close to the commonly applied conventions for tolerable error (i.e., 5%). Moreover, we found that the *separation reliabilities* (for details, see

Wright & Masters, 1982) for the five scales (.74 to .88) were comparable to the reliability estimates found for conventional attitude measures in psychology.

Table AI-1

*Formal scale features of the five environmental attitude Rasch scales.*

| | EA-50 | EA-48 | EA-98 | EA-49A | EA-49B |
|---|---|---|---|---|---|
| ITEM FIT STATISTICS: | | | | | |
| *M(MS)* | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| *SD(MS)* | 0.05 | 0.06 | 0.06 | 0.06 | 0.06 |
| *Minimum (MS)* | 0.89 | 0.84 | 0.87 | 0.89 | 0.87 |
| *Maximum (MS)* | 1.10 | 1.18 | 1.16 | 1.13 | 1.17 |
| PERSON FIT STATISTICS: | | | | | |
| *M(MS)* | 1.01 | 1.00 | 1.00 | 1.00 | 1.00 |
| *SD(MS)* | 0.26 | 0.22 | 0.20 | 0.25 | 0.25 |
| *% people with poor fit ($t \geq 1.96$)* | 4.57% | 6.48% | 9.28% | 6.61% | 6.35% |
| SCALE FEATURES: | | | | | |
| *Separation reliability* | .74 | .83 | .88 | .76 | .81 |
| *Kurtosis* | 1.69 | 0.28 | 0.35 | 0.40 | 0.54 |
| *Skewness* | 0.45 | 0.33 | 0.29 | 0.25 | 0.27 |

*Note*. EA-50 was based on 50 behavioral self-reports; EA-48 was based on 48 evaluative statements; EA-98 was a combination of 50 self-reports and 48 evaluative statements; and EA-49A and EA-49B were two versions of 49 items each consisting of 25 self-report items and 24 evaluative statements from the EA-98 scale.

Although all five scale calibrations revealed generally reasonable and comparable results, the specific statistical fit of the five instruments nevertheless revealed some differences. As expected, the all-inclusive (i.e., the EA-98) and thus the longest attitude scale turned out to be the best with respect to its ability to separate people (i.e., in terms of reliability); it was also the most sensitive for detecting poorly fitting people. Jointly, these findings indicate that more information is better for distinguishing individuals and for recognizing even small model inaccuracies. We also found that when attitude measurement is exclusively based on evaluative statements (EA-48), it is a bit more fallible than when it is exclusively based on behavioral self-reports (EA-50: see Table AI-1).

## AI-References

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

DeCoster, J., Iselin, A.-M. R., & Gallucci, M. (2009). A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods, 14,* 349-366.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

Kaiser, F. G., Hartig, T., Brügger, A., & Duvier, C. (2013). Environmental protection and nature as distinct attitudinal objects: An application of the Campbell paradigm. *Environment and Behavior, 45,* 369-398.

Kaiser, F. G., & Wilson, M. (2000). Assessing people's general ecological behavior: A cross-cultural measure. *Journal of Applied Social Psychology, 30,* 952-978.

Kaiser, F. G., & Wilson, M. (2004). Goal-directed conservation behavior: The specific composition of a general performance. *Personality and Individual Differences, 36,* 1531-1544.

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scales? Study 1: Reliability and validity. *Educational and Psychological Measurement, 31,* 657-674.

Peabody, D. (1962). Two components in bipolar scales: Direction and extension. *Psychological Review, 69*, 65-73.

Wright, B. D., Linacre, J. M., Gustafson, J.-E., & Martin-Löf, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement.* Chicago: MESA.

## APPENDIX II

Table AII-1

*All 45 difference-ratios combinations of five person estimates assessed with two environmental attitude scales—EA-49A and EA-49B—either in logits- or in z-scores.*

| | Ratios of differences | | | Ratios of differences | | |
|---|---|---|---|---|---|---|
| | *EA-49A in logits* | *EA-49B in logits* | % | *EA-49A in z-scores* | *EA-49B in z-scores* | % |
| 1. | 0.84 | 0.83 | 1.17 % | 0.56 | 0.57 | 2.82 % |
| 2. | 0.64 | 0.63 | 1.02 % | 0.50 | 0.50 | 0.00 % |
| 3. | 0.47 | 0.46 | 1.49 % | 0.36 | 0.36 | 1.80 % |
| 4. | 5.14 | 4.79 | 6.99 % | 1.25 | 1.33 | 6.45 % |
| 5. | 1.77 | 1.72 | 2.79 % | 1.00 | 1.00 | 0.00 % |
| 6. | 0.88 | 0.86 | 2.79 % | 0.56 | 0.57 | 2.82 % |
| 7. | 2.69 | 2.68 | 0.52 % | 5.00 | 4.00 | 22.22 % |
| 8. | 1.07 | 1.05 | 1.90 % | 1.00 | 1.00 | 0.00 % |
| 9. | 1.77 | 1.72 | 2.79 % | 1.25 | 1.33 | 6.45 % |
| 10. | 0.76 | 0.76 | 0.15 % | 0.90 | 0.88 | 2.82 % |
| 11. | 0.56 | 0.56 | 0.32 % | 0.64 | 0.64 | 1.02 % |
| 12. | 6.14 | 5.79 | 5.82 % | 2.25 | 2.33 | 3.64 % |
| 13. | 2.11 | 2.08 | 1.62 % | 1.80 | 1.75 | 2.82 % |
| 14. | 1.05 | 1.04 | 1.62 % | 1.00 | 1.00 | 0.00 % |
| 15. | 3.21 | 3.24 | 0.65 % | 9.00 | 7.00 | 25.00 % |
| 16. | 1.27 | 1.26 | 0.73 % | 1.80 | 1.75 | 2.82 % |
| 17. | 2.11 | 2.08 | 1.62 % | 2.25 | 2.33 | 3.64 % |
| 18. | 0.73 | 0.73 | 0.47 % | 0.71 | 0.73 | 1.80 % |
| 19. | 8.05 | 7.58 | 5.97 % | 2.50 | 2.67 | 6.45 % |
| 20. | 2.77 | 2.72 | 1.77 % | 2.00 | 2.00 | 0.00 % |
| 21. | 1.38 | 1.36 | 1.77 % | 1.11 | 1.14 | 2.82 % |
| 22. | 4.21 | 4.24 | 0.50 % | 10.00 | 8.00 | 22.22 % |
| 23. | 1.67 | 1.66 | 0.88 % | 2.00 | 2.00 | 0.00 % |
| 24. | 2.77 | 2.72 | 1.77 % | 2.50 | 2.67 | 6.45 % |
| 25. | 10.95 | 10.37 | 5.50 % | 3.50 | 3.67 | 4.65 % |
| 26. | 3.77 | 3.72 | 1.30 % | 2.80 | 2.75 | 1.80 % |
| 27. | 1.88 | 1.86 | 1.30 % | 1.56 | 1.57 | 1.02 % |
| 28. | 5.74 | 5.79 | 0.97 % | 14.00 | 11.00 | 24.00 % |
| 29. | 2.27 | 2.26 | 0.41 % | 2.80 | 2.75 | 1.80 % |
| 30. | 3.77 | 3.72 | 1.30 % | 3.50 | 3.67 | 4.65 % |

continued

| | Ratios of differences | | | Ratios of differences | | |
|---|---|---|---|---|---|---|
| | *EA-49A in logits* | *EA-49B in logits* | % | *EA-49A in z-scores* | *EA-49B in z-scores* | % |
| 31. | 0.34 | 0.36 | 4.20 % | 0.80 | 0.75 | 6.45 % |
| 32. | 0.17 | 0.18 | 4.20 % | 0.44 | 0.43 | 3.64 % |
| 33. | 0.52 | 0.56 | 6.47 % | 4.00 | 3.00 | 28.57 % |
| 34. | 0.21 | 0.22 | 5.09 % | 0.80 | 0.75 | 6.45 % |
| 35. | 0.34 | 0.36 | 4.20 % | 1.00 | 1.00 | 0.00 % |
| 36. | 0.50 | 0.50 | 0.00 % | 0.56 | 0.57 | 2.82 % |
| 37. | 1.52 | 1.56 | 2.27 % | 5.00 | 4.00 | 22.22 % |
| 38. | 0.60 | 0.61 | 0.89 % | 1.00 | 1.00 | 0.00 % |
| 39. | 1.00 | 1.00 | 0.00 % | 1.25 | 1.33 | 6.45 % |
| 40. | 3.05 | 3.12 | 2.27 % | 9.00 | 7.00 | 25.00 % |
| 41. | 1.21 | 1.22 | 0.89 % | 1.80 | 1.75 | 2.82 % |
| 42. | 2.00 | 2.00 | 0.00 % | 2.25 | 2.33 | 3.64 % |
| 43. | 0.40 | 0.39 | 1.38 % | 0.20 | 0.25 | 22.22 % |
| 44. | 0.66 | 0.64 | 2.27 % | 0.25 | 0.33 | 28.57 % |
| 45. | 1.66 | 1.64 | 0.89 % | 1.25 | 1.33 | 6.45 % |
| | | mean: | 2.07% | | mean: | 7.27% |

*Note*. EA-49A and EA-49B reflect the difference ratios for the environmental attitude measures A and B either in logits or in z-scores. The % column stands for the relative difference between the two ratios (relative to the average of the two ratios).