Dear Editors,

Environmental psychologists are interested in a diverse array of constructs, including but not limited to environmental behaviors (Steg & Vlek, 2009; Urban & Braun Kohlová, 2022), environmental attitudes (Milfont & Duckitt, 2004; Wyss et al., 2022), environmental concern (Schultz, 2001), environmental beliefs (Dunlap & Van Liere, 1978; Rosa et al., 2022), and connection to nature (Coughlan et al., 2022; Ives et al., 2017). Valid and reliable measurement of such constructs is essential for research and practice. Invalid measurement can lead to misleading inferences. While most researchers are aware of this, recent work has revealed that many measures have critical limitations or are used inappropriately (Hawcroft & Milfont, 2010; Mokkink et al., 2018; Perrin & Benassi, 2009; Rosa et al., 2022; Stallwood et al., 2021; Terwee et al., 2018). In this letter, we use examples of frequently utilized and widely cited scales to illustrate four key challenges and corresponding recommendations regarding measurement in the field of environmental psychology.

Clarify construct definition and operationalization. In psychological research, the construct(s) of interest should be defined in sufficient detail to enable the development of a measure that covers the particular target construct. The popular New Ecological Paradigm (NEP) offers one example of an ambiguously defined construct (Dunlap et al., 2000; Dunlap & Van Liere, 1978). Developers of the NEP scale acknowledged uncertainty regarding what the scale was purported to measure, suggesting the NEP construct was “somewhat amorphous” (Dunlap et al., 2000, p. 429), including beliefs related to the balance of nature, the existence of ecological limits, and humans’ role as a part of nature. It is difficult to know how well a scale assesses the construct it aims to measure without sufficient information about what that construct is (Rosa et al., 2022). In the case of the NEP, we cannot evaluate whether all dimensions of this amorphous construct are effectively covered by the items on the NEP scale(s). Whereas there is no consensus regarding what constitutes a sufficient characterization of a construct, it generally requires a definition as well as specific examples of what a construct is and what it is not. Ideally, construct definition also considers theoretical relations with related constructs, for example, within a nomological network (Cortina et al., 2020; Flake & Fried, 2020). A further definition of the NEP might therefore include specifying all aspects of this construct and differentiating the NEP from other constructs like environmental attitudes, environmental beliefs, and connectedness to nature (Rosa et al., 2022).

Consider face validity and construct coverage. After describing the construct(s) of interest in sufficient detail, researchers should show how the content of their measure matches the content of the construct. For example, the Connectedness to Nature Scale (CNS, Mayer & Frantz, 2004, p. 593) was designed to assess “individuals’ trait levels of feeling emotionally connected to the natural world.” Mayer and Frantz’s (2004) description of the construct led other researchers in the field to question whether the CNS actually assesses feelings of emotional connection rather than related constructs such as beliefs about an individual’s dependence on nature (Pasca et al., 2017; Perrin & Benassi, 2009). To investigate whether a measure matches the construct under investigation, scale developers can gather expert opinions regarding the relevance of each scale item and qualitatively describe how each item’s content is related to the construct of interest (Terwee et al., 2018). For example, experts in connectedness to nature can be provided with the construct definition and asked to indicate whether they think all the items of the CNS are relevant for this construct and if the items cover all aspects of connectedness to nature. Possible conclusions from such an evaluation could be that a construct needs to be better defined, that specific items are not related to the construct, or that items collectively fail to capture some important dimensions of the construct.

Examine interpretation of item(s) and response processes in diverse participants. Valid measurement of a construct typically requires that a measure is interpreted by the target population as intended by the scale developers (AERA & NCME, 2014; Peterson et al., 2017). If respondents and scale developers differ in their interpretations, or if respondents differ from each other in their interpretations, or if a scale developer does not fully understand how respondents interpret items, measurement problems may ensue (AERA & NCME, 2014). Taking the NEP scale for children as an example, the item “People must still obey the laws of nature” is used to represent the idea of “human exemptionalism” within the broader NEP (Dunlap et al., 2000; Manoli et al., 2007). However, if children have a different understanding of the expression “laws of nature” than the one expected by the scale developers, the idea of “human exemptionalism” may be inaccurately represented (Harrison, 2020; Rosa et al., 2022). Similarly, varying conceptualizations of “nature” among children (Collado et al., 2016; Larson et al., 2011) and adults (Muhar et al., 2018) might lead researchers to draw inaccurate inferences related to items that include the word “nature.” Cognitive
interviews, in which members of the target population respond to an instrument while expressing their thoughts aloud, can help researchers address this challenge through a better understanding of diverse participants’ response processes (Peterson et al., 2017). For example, Harrison (2020) interviewed children and adolescents to understand their thought processes when responding to items on the NEP scale for children, and Clayton et al. (2021) engaged in discussions and workshops across five countries to characterize potential cultural variations in children, and Clayton et al. (2021) engaged in discussions and workshops across five countries to characterize potential cultural variations in participants’ interpretation of the Environmental Identity Scale.

Align theoretical and statistical models. Scale developers nearly universally rely on reflective latent variable models (LVMs), such as exploratory and confirmatory factors models, to evaluate measures and justify changes (e.g., rephrasing or dropping items). These models impose assumptions on data that may require more consideration in the literature. One example is the common cause theory. In substantive language, this means that responses on items $x_1, x_2, x_3, \ldots x_n$ are caused (and only caused) by a common construct $Y$ (Van Bork et al., 2017). For instance, because of the assumption that the personality trait extraversion influences how people behave (McCrae & Costa, 1995), this trait is measured with questions about certain types of behavior. This implies that shared variance among items is caused by the underlying construct, and variance that is not shared is measurement error (for details, see Fried, 2020; Rhemtulla et al., 2020; Van Bork et al., 2017). However, this assumption may not always be accurate, because different causal models can lead to shared variance among items $x_1, x_2, x_3, \ldots x_n$ other than a common factor $Y$: this might occur when one item $x_1$ causes another item $x_2$. For example, consider the following items from Nisbet and Zelenksi’s (2013) short version of the Nature Relatedness Scale (Y): “I feel very connected to all living things and the earth” ($x_1$) and “My ideal vacation spot would be a remote, wilderness area” ($x_2$). These two items may not be statistically independent given $Y$, because feeling connected with nature ($x_1$) might stimulate people’s preferences for activities in remote natural areas ($x_2$), which could in turn reinforce $x_1$ (Barrable & Booth, 2020; Rosa et al., 2020). Though the LVM can be used without a causal theory, the utility of this model in these cases is limited (Van Bork et al., 2017). Therefore, we believe that researchers should provide an explanation as to why the LVM, when utilized, is best suited for representing a focal construct at hand. Some authors have considered this. As an example, Kaiser and Lange (2021) propose that people’s environmental attitudes can motivate engagement in environmental behaviors, so these attitudes should (at least in part) explain the shared variance among self-reported environmental behaviors. We recommend readers consult other studies for further information regarding measurement model selection and use (Fried, 2020; Rhemtulla et al., 2020; Schmittmann et al., 2013; Van Bork et al., 2017). These four measurement challenges are not unique to environmental psychology, and we do not see specific characteristics that make measurement harder in our field than in other areas of psychology. We chose to focus on these particular challenges because they are very common in environmental psychology literature, and they can, in principle, be resolved with greater attention to the content of a construct and measure. Our aim in presenting these challenges is not to criticize decades of work in the field that has focused on the development of measures and assessment of constructs. Rather, we hope that outlining challenges and future directions will give rise to necessary discussions about measurement practices in environmental psychology that will help the field to better achieve its goals. We recommend that researchers consult other sources for further guidance. We refer here to developing a construct (Cortina et al., 2020; Fluke & Fried, 2020; Rhemtulla et al., 2020), validate a measure (AERA & NCME, 2014; Cook et al., 2015; Kane, 2013; Mokkink et al., 2018; Terwee et al., 2018) and report a validation study (Fluke & Fried, 2020; Gagnier et al., 2021). These sources offer guidance for addressing challenges not only during the development of new scales but also when utilizing previously developed scales. It is ultimately the researcher’s responsibility to ensure that a chosen measure is valid for the specific use to which it is being applied (AERA & NCME, 2014; Kane, 2013).

Hopefully, our suggestions will help to develop the improvement and use of measures in environmental psychology and inspire future research on this important topic.

Funding

This study was partially funded by the Spanish Ministry of Science, Education, and Universities (PGC2018-095502-B-I00) and by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Declaration of competing interest

We have no conflict of interest to disclose.

References


