International Journal of Early Childhood Environmental Education

ISSN: 2331-0464 (online)

Psychometric Validation of a Game-based Testing Instrument to Measure Preschool Children's Environmental Knowledge and Connection to Nature

Jessica MacKeen
Dalhousie University, Canada
Tarah Wright
Dalhousie University, Canada
Daniel Séguin
Mount Saint Vincent University, Canada
Heather Cray
Dalhousie University, Canada

Submitted April 26, 2022; Accepted November 3, 2022

ABSTRACT

In this study, we use face and content validity to determine whether a modified game-based testing instrument is appropriate and relevant for quantifying preschool children's emotional, cognitive, and attitudinal affinity with nature. Six environmental psychology experts completed a questionnaire and subsequent interviews with three of them provided insight into whether the features of the tool can measure a child's affinity with nature. Interrater agreement and content validity indexes establish whether the instrument meets four established validation criteria: clarity; ease of use; appropriateness; and relevancy. Results indicate that the modified device does not measure the original core concept: affinity with nature, and that face and content validation outcomes do not yield acceptable outputs. As a result, the foundation and key characteristics of the instrument were revised to enhance its ability to produce trustworthy results and more accurately measure the revised core concept of connection to nature and environmental knowledge.

***Keywords*:** connection to nature, environmental knowledge, psychometric evaluation, validity, nature

Exposure to the natural world is known to be one of the most essential components during the early stages of a child's physical, attitudinal, intellectual, and moral development (Kahn & Kellert, 2002). Studies indicate that spending time outdoors correlates with increased physical activity, leading to many health benefits such as building and maintaining healthy bones and muscles, and reducing risk of chronic diseases, depression, and anxiety (McCurdy *et al.*, 2010; Mygind *et al.*, 2019). One's emotional and attitudinal connection to nature (CTN) is largely influenced by positive and frequent experiences in outdoor environments (Clayton & Opotow, 2003). These exposures have been proven to increase the development of pro-environmental attitudes, knowledge, and beliefs as an adult (Chipeniuk, 1995; Ewert *et al.*, 2005; Rickinson, 2001), and the probability of conservation behaviours and attitudes later in life (Dresner *et al.*, 2015; Zhang *et al.*, 2014). Other studies in the field have found that time spent in nature enhances children's development of imagination, creativity, and problem-solving skills and their overall connection to nature (Chawla, 2015; MacKeen & Wright, 2020; Malone & Tranter, 2003; Omidvar, 2018; Omidvar *et al.*, 2019).

Understanding the variety of relationships young children experience with nature is critical for enhancing education systems, children's relationships with others, and their physical and psychological development (Braus & Milligan-Toffler, 2018; Kahn & Kellert, 2002). Further, by studying and supporting children and their connections to nature,

we gain a sense of how caring for the environment can be fostered at these impressionable ages. Despite the growing amount of literature exploring children's CTN, information and measurement of the impacts nature has on children during various stages of development are sparse, and gaps remain, such as whether young children can form deep connections with nature and how direct or indirect contact with nature plays a role in their relationships with the natural world (Kahn & Kellert, 2002).

Connection to nature looks different for each individual, especially young children. Studies have started to acknowledge that there is no single right way to interact with nature, and as a result, there is no single effective way for researchers to study human nature connection (Braus & Milligan-Toffler, 2018). Because of this variety, researchers have measured these relationships using an assortment of constructs. The construct is a fundamental concept a researcher intends to explore through methods such as psychological testing instruments (Cronbach & Meehl, 1955). Constructs related to children's connection to nature have been defined as nature relatedness (Nisbet et al., 2009), the inclusion of nature in self (Schultz, 2002), eco-affinity and eco-awareness (Larson et al., 2011), and nature connection (Braus & Milligan-Toffler, 2018; J. C.-H. Cheng & Monroe, 2012; Kellert, 2012; Mayer & Frantz, 2004). While these concepts are not the same, they do encompass the affective, cognitive, and experiential aspects associated with measuring nature relationships (Barrable & Booth, 2020; Tam, 2013). The debate remains on which method is more effective at measuring psychological attributes; aggregating similar concepts into a broader construct or narrowing the concepts to be more specific (Tam, 2013). Nevertheless, precise definitions are essential to ensure that a construct is feasible for measuring the targeted concepts.

A select number of psychological testing instruments strive to explore children's different relationships with nature (Giusti *et al.*, 2014; Larson *et al.*, 2011; MacKeen & Wright, 2020; Mayer & Frantz, 2004; Nisbet *et al.*, 2009). Both Larson *et al.*, (2011) and Nisbet *et al.*, (2009) studied the affective and cognitive aspects of an individual's CTN via survey-based scales. Mayer and Frantz (2004) utilized an instrument built to measure one's emotional connection with nature and suggested that this connection is an essential predictor of ecological behaviour and personal well-being. While these devices and others alike have been psychometrically evaluated for validity and reliability, there is no evidence of a trustworthy game-based testing instrument built for measuring preschool children's CTN and environmental knowledge. The only known tool of this kind was conceptualized in 2012 and created in 2014, where Giusti *et al.* developed a measure that was said to assess 5-year-old children's emotional, cognitive, and attitudinal affinity with nature.

Giusti *et al.* (2014) found that students in local Reggio-Emilia (nature-based philosophy) schools had increased bioaffinity (CTN) and the device yielded strong internal consistency reliability results. Using the Giusti *et al.* (2014) tool, Omidvar *et al.* (2019) conducted a similar study in Nova Scotia, Canada, with 3-5-year-old children attending a Reggio-Emilia preschool but results indicate their affinity with nature was much weaker than hypothesized. As a result, Omidvar *et al.* (2019) recommended that further studies determine the appropriateness of the Giusti *et al.* (2014) measure for younger Canadian children. As a result, MacKeen & Wright (2020) modified the Giusti *et al.* (2014) tool to be more culturally, geographically, and developmentally relevant for young children in Canada, and then pilot tested the revised version. Modifications primarily included changes to the game design and the content (i.e., pictures and language). Results showed that the revisions effectively enhanced the children's understanding of the game's testing for a Canadian context. However, for psychological instruments to be considered trustworthy and produce generalizable results, they must undergo psychometric assessment (DeVellis & Thorpe, 2021). While Giusti *et al.*, (2014) established initial reliability results for the original version of the testing instrument, neither the 2014 or modified 2020 version have undergone any type of validity assessment.

Validation in Research

Validity is a vital step in producing effective and quality research. Validation procedures emerged as a means to determine the degree to which a psychological or educational test evaluates what it proposes to measure (Sireci, 1998). Throughout its evolution, new versions of validity testing have emerged to aid in assessing different characteristics of instruments, such as face, content, criterion-related, construct, external and others (Cohen *et al.*, 2002). For example, construct validity is used to investigate the foundational concept that the device is built to

measure, and external validity looks at the degree to which the results are generalizable to greater populations and locations (Cohen *et al.*, 2002). Given the variety of available tests, researchers utilize the type that is most appropriate for their device and field. Limpo *et al.* (2020), examine construct, convergent/discriminant, and predictive validity for an instrument that measures Portuguese third graders' self-efficacy, and motivations for writing and story-telling. These methods were used to examine factorial structure and invariance across two independent samples of third graders, which were used as a procedure for assessing text quality across different genres and grade levels (Limpo *et al.*, 2020). Another study investigated the concurrent validity of a tool that measures students' science attitudes, pro-environmental behaviour, climate change attitudes, and knowledge (Dijkstra & Goedhart, 2012). An expert panel was asked to assess these variables amongst different groups within the study population (Dijkstra & Goedhart, 2012). However, Giusti *et al.*, (2014) did not establish any form of validity for the original Games Testing for Emotional Affinity, Cognitive Affinity, and Attitudinal Affinity with the Biosphere instrument.

Preliminary validity testing involves looking at the basic foundations of the tool, which could include but is not limited to assessing face, content, and construct validity criteria. Face and content validity analyses were chosen for this study because they assess the whether the premise of the instrument is effective for measuring the foundational concept. Face validity seeks to determine whether the tool's items are sensible, appropriate, and relevant to the participant audience (Connell *et al.*, 2018). This method relies on knowledgeable experts reviewing the suitability of the items within an instrument pertaining to the measured psychological criteria (Connell *et al.*, 2018; Holden, 2010). Slater *et al.* (2009) recruited an expert panel of 6 individuals comprised of nurses and psychologists with experience in clinical practice, research, and survey design due to their instrument being focused on measuring nurses' working environments. Similarly, a study by Piredda *et al.* (2017) utilized a panel of 6 experts to assess an instrument's items for clarity and appropriateness of a tool that sought to measure nurses' caring behaviour in Italian acute care settings. Both studies used a survey, semi-structured or structured interviews with the experts to gather the information, which was then transcribed and analyzed for common consensus (Njelesani *et al.*, 2020; Piredda *et al.*, 2017; Slater *et al.*, 2009).

Throughout the literature, face validity is often paired with content validity (Connell *et al.*, 2018; Krause, 2012; Njelesani *et al.*, 2020; Piredda *et al.*, 2017; Slater *et al.*, 2009). Haynes *et al.* (1995) defines content validity as "the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose" (p. 238). The construct refers to the concept that the tool intends to measure, which in the case of the MacKeen and Wright (2020) version of the tool is the cognitive, emotional, and attitudinal bioaffinity (CTN) of preschoolers (Haynes *et al.*, 1995). Content validity can be established via quantitative, qualitative, or a mixed-methods approach, and primarily relies on the following criteria: establish a clear definition of the construct or concept the tool aims to measure, gather expert opinions for the items within the tool using formalized scaling procedures, and examine the proportional representations of the items to determine whether the tool is interpreted in a way that reflects the construct or concept (Haynes *et al.*, 1995). Silva *et al.*, (2020) followed these steps by recruiting 10 expert raters and asking them to rate the items as relevant or not relevant on a formalized scale during the interview process. Validity allows the researcher to determine whether the items in an instrument are useful for measuring the targeted concept and identify areas that could be enhanced to better measure the construct (Roberts et al., 2006). It is apparent that validity is an important step in creating a sound psychological instrument, yet both the original (Giusti *et al.*, 2014) and modified (MacKeen and Wright, 2020) versions of the game-based testing instrument lack validation. Establishing validity of the MacKeen and Wright 2020 instrument will highlight whether the instrument and its components are useful for measuring young children's environmental knowledge and connection to nature.

An apparent gap exists concerning a recognised valid connection to nature and environmental knowledge scale using a game-based testing strategy, specifically for young children. Task-based learning can help children learn more effectively due to multiple parts of the brain being engaged, thus allowing the child to stay focused (Buyukkarci, 2009; Littlewood, 2004). The game-based instrument includes high-level concepts that may not have been previously introduced or particularly interesting to a young audience. The game-based format increases the chance of engagement and the probability of the information having a lasting impression on the child (Buyukkarci, 2009). Though there is substantial literature concerning the validity of psychological CTN related testing tools (i.e.,

questionnaires and surveys), there is a lack of literature exploring a testing tool similar to the modified Giusti *et al.* (2014) games testing tool. Existing literature focuses on developing individual tools that measure different human behaviours, and some of these studies do investigate connection to nature (Cheng & Monroe, 2012; Mayer & Frantz, 2004; Nisbet *et al.*, 2009). However, none of the existing research seeks to evaluate this criterion through a games testing tool.

The current research seeks to establish preliminary validity assessments for the MacKeen and Wright 2020 modified game-based testing instrument for measuring Canadian preschool children's CTN. Specifically, face and content validation will be explored through applying a mixed-methods approach by testing whether the newly modified 2020 tool is more valid than the original 2014 tool in terms of face (ability) and content. Appraisal of the face and content psychometrics includes the use of an expert panel who are asked to review and be interviewed about the ability and suitability of the modified instrument based on four criteria: clarity, ease of use, appropriateness, and relevancy (Connell *et al.*, 2018; Haynes *et al.*, 1995; Holden, 2010). These two types of substantiation provide a foundation for other forms of validity and psychometric evaluation to take place as the tool moves towards becoming trustworthy and generalizable. Further, by initiating the validation of the device, it has the potential to have an interdisciplinary impact and contribute to various research sectors by allowing future studies to further determine whether time spent in nature and environmentally-focused school situations can increase a preschooler's CTN and environmental knowledge.

Methods

Before a tool can undergo psychometric reliability testing, validity must be evaluated as it determines whether the criterion within the tool is useful for measuring the intent of the tool. Once validity has been assessed, the tool can be used in a chain of pilot studies where reliability can be evaluated. In this study, a mixed-methods approach is used to assess the face and content criteria for the MacKeen and Wright (2020) modified CTN game-based testing instrument. Both face and content validity are confirmed via an expert (i.e., practitioners and academics) panel providing feedback and insight through a questionnaire and follow-up semi-structured online interviews. Ethics approval was granted before data collection commenced via the Dalhousie University Research Ethics board. For this research, the use of face and content criteria served the purpose of establishing a foundational understanding of the intent and ability of the characteristics in the modified instrument used to measure CTN and environmental knowledge. Further, these forms of validation are critical steppingstones for other types of validity to be later established.

The tool

The game-based testing instrument used in this research was initially designed by Giusti *et al.* (2014) with games that were meant to be played on a standard size piece of paper (8.5 x 11 inches); however, the varying types of images (cartoon and real pictures), and much of the language was not appropriate for a Canadian context. As such, the tool was modified by MacKeen and Wright (2020), primarily including changes to the game design, pictures and language (please note that a full copy of the instrument is available by contacting the authors or by visiting this link). The modified instrument that was used for this portion of the study contained six unique task-oriented games that utilize monochromatic photos of real-life items (i.e., a real photo of a tree or people cleaning up a beach) and culturally and developmentally appropriate language for Canadian preschoolers.

The first game (game 1a, see Table 1) seeks to explore children's environmental sensitivity by using a sorting game that asks the child whether certain inanimate objects and animals can get hurt. It includes cut-outs of particular photos (i.e., a tree and a bird), and the child is asked to sort the pictures into 'yes' or 'no' bins based on the question posed. The second game (game 1b) is also used to test environmental sensitivity by employing a game of happy and sad faces, where there is a board of pictures (i.e., water pollution and planting a tree) and cut out happy and sad faces. The children are then asked for each picture on the board, whether they want to associate a happy or sad face with that picture.

The next two games gauge children's environmental awareness. Game 2a is a matching game that includes a board of pictures and cut-outs. The board displays photos such as eggs and paper, and then for each cut-out, they are asked to match them with the corresponding ecosystem service (i.e., chicken and wood). The fourth game (game 2b) is completed in two parts. First, the child is asked to verbally explain their definition of particular pollutants (i.e., water pollution). Secondly, they are shown a list of the pollutants and asked whether they can hurt things, such as cars and animals.

The final two games aim to measure children's environmental preferences and are played verbally; they both use the same board of photos depicting physical places that children play (i.e., a backyard or playground). The fifth game (game 3a) asks the participants about their favorite places to play and why, and game 3b asks the participants about their least favorite places to play and why. The total amount of time needed to perform the testing is an average of 15.25 minutes (MacKeen and Wright, 2020).

Table 1
List of games and the concepts they intend to measure and game style used.

Game	Concept Measured and Game Style
Game 1A	• Children's environmental sensitivity • Sorting game
Game 1B	• Children's environmental sensitivity • Matching game using happy and sad faces
Game 2A	• Children's environmental awareness • Matching game of ecological services
Game 2B	• Children' environmental awareness • Verbal response and sorting game
Game 3A and 3B	• Children's environmental preferences • Participants choice of photo and verbal response

Data collection

Data for both face and content validity was obtained through a questionnaire and, when appropriate, follow-up interviews with a group of experts within the field of environmental psychology, early childhood environmental education, and connection to nature and bioaffinity. Recruitment of the expert panel (n=6) was carried out through a non-probabilistic and purposive sampling technique, focusing on a combination of stakeholder and criterion sampling that allowed the researchers to identify and interview significant stakeholders who are knowledgeable about CTN and the creation, use, and evaluation of psychological testing instruments. The chosen experts have participated in creating and evaluating psychological testing instruments designed to measure children's psychological attributes related to nature connection. Therefore, they possess an understanding of the theory and methods used to assess this particular type of instrument. Participants were invited via e-mail correspondence that included information about the study, tasks, and a consent form, as well as preemptive consent for a follow up interview and permission to digitally record the interview. When the specialists agreed to participate, the study's questionnaire was sent out for completion.

Questionnaire construction

The purpose of the questionnaire was to formally explore the four criteria used to assess face (clarity, ease of use, and appropriateness) and content (relevancy) validity. The questionnaire posed a number of Likert-style questions to assess the four criteria for the instructions, pictures, and language used in the modified tool. For each of the three components of the modified tool, a five-point scale (i.e., 1 = very unclear, 2 = unclear, 3 = neutral/undecided, 4 = clear, 5 = very clear) was composed to assess the individual qualities (clarity, ease of use, appropriateness, and relevancy). Clarity was used to investigate the comprehensiveness of the intent and contents, ease of use and

appropriateness contributed to whether the content was considered culturally and/or developmentally suitable for Canadian preschoolers, and relevancy was used to establish if the items in the tool were representative of the targeted construct (the foundational concept the tool was built to measure – in this case, connection to nature) (Piredda *et al.*, 2017; Polit & Beck, 2009; Silva *et al.*, 2020). The scale provided an extensive amount of rich information concerning each section of the tool, including the information provided before starting the game's testing, each of the six games, and the debrief section.

Interviews

Once the questionnaires were complete, the face validity analysis determined whether semi-structured follow-up interviews were necessary to investigate poor results. Interviews were conducted through the online platform Zoom as it is considered the most popular video conferencing app and can provide password protection for the meeting, the ability to lock up the meeting (mitigating against unwanted users joining the meeting), and individual privacy controls (Singh & Awasthi, 2020). Of the six experts taking part in this validity testing of the modified tool, three were interviewed to gain further insight into the answers they provided in their questionnaires. The survey results indicated what areas of the tool were lacking in terms of validity but there was no space for the experts to provide reasoning. The scores for three of the experts highlighted that they were satisfied with the items in the tool and how they operate to measure the construct of the instrument. However, the scores for the other three experts specified that the instrument needed further modification. Interviews were chosen to explore their answers and gain insight for how to enhance the instrument and its ability to measure children's connection to nature. A rough interview guide was created to cater to each of the three interviews because of the differing questionnaire results. Therefore, each interview produced varying perceptions about the expert's concerns within the tool. These varying outlooks were then compiled to examine for emerging patterns and to determine the best way to modify specific segments in the instrument.

Data Analysis

Just as there are various ways to evaluate the validity of a testing instrument, there are also a variety of methods used to analyze the data. Previous work on this subject suggests that the first three criteria (clarity, ease of use, and appropriateness) can be calculated together as they examine the intent and suitability of the items within the tool, which indicates if the face value of the tool is adequate (Piredda *et al.*, 2017; Silva *et al.*, 2020). In comparison, the criterion relevancy looks at whether the items are pertinent for measuring the instrument's construct. As a result, analyses differ with face validity utilizing interrater reliability correlation analysis and content validity using the item and scale content validity index calculations (Polit et al., 2007; H. E. A. Tinsley & Weiss, 2000).

Face Validity

To assess the face validity criteria collected via the questionnaire, interrater agreement analysis and the intra-class correlation coefficient calculation was used. Interrater agreement indicates the degree to which the ranking amongst a group of raters is the same or, more specifically, determines the strength between two or more raters (Tinsley & Weiss, 2000). With a five-point scale, high interrater scores suggest that the experts value the features of the tool (i.e., pictures) with a consistent rating (Tinsley & Weiss, 2000). We chose the intra-class correlation coefficient (ICC) for this research due to its widespread use in social science literature and the number of expert raters (n=6) recruited for this study (Bajpai *et al.*, 2015; Bartko, 1966). ICC results range on a spectrum between zero to one, with 0.70 being considered an acceptable level of agreeance between the expert raters (Tinsley and Weiss, 2000).

Before calculations could commence, the expert scores gathered via the questionnaire were converted to a zero (expert ratings equaling to three or below) or a one (expert ratings equaling to four or five). Then, the ICC and other descriptive statistics (mean, variance, and 95% confidence intervals) were calculated individually for clarity, ease of use, and appropriateness via the statistical package SPSS. The confidence interval (CI) results are shared because they provide a deeper understanding of the relationship under scrutiny (Kallogjeri *et al.*, 2019). This data showcases the variability, which aids in getting a broader picture of insight about certain outcomes (Kallogjeri *et al.*, 2019).

Content Validity

The item and scale-level content validity index (CVI) is used to quantify content validity for the multi-item modified scale based on expert ratings of relevance or their average of agreeance (Polit *et al.*, 2007). Calculating CVI is an essential step in validating the foundations of an instrument and ensuring high-quality content (Polit *et al.*, 2007; Shi *et al.*, 2012). All CVI calculations were conducted in Microsoft Excel.

Before calculating CVI, the scores from the Likert-style questions were converted to a zero (expert ratings equaling to three or below) or a one (expert ratings equaling to four or five) to conform to the range used to assess CVI. Once the scores were converted, the first calculation was the item-level content validity (I-CVI), which was used to determine the CVI for each individual component (i.e., game 1a, game 1b, etc.). Then, the I-CVI was used to compute the average I-CVI across the items, known as the scale-level content validity average (S-CVI) (Polit *et al.*, 2007). More specifically, the (S-CVI) looks at all of the components (i.e., all six games and the instructional sections in the tool) being assessed by the expert raters. The analysis includes the I-CVI and S-CVI average for the pictures, language, and instructions for each of the games within the modified tool. When evaluating new testing instruments, a typical limit of acceptability for the S-CVI average is a score of 0.80 or higher (Polit *et al.*, 2007; Shi *et al.*, 2012).

Results

The following section showcases the results from the face and content validity analyses and utilizes the interview outcomes to help support and illustrate them. The four components of the instrument (instructions, recommendations, language, and pictures) that the experts analyzed are used to guide the results and discussion to follow. Finally, it is essential to note that the outcomes from the validity analysis facilitated modifications to the MacKeen and Wright 2020 instrument that resulted in a new version of the instrument dated 2022.

Face Validity

The validation scores for the questionnaire's face (ability) assessment are categorized into four sections: recommendations, instructions, pictures, and language. Within these sections, readers will find explanations about each of the three-face validity criteria clarity, ease of use, and appropriateness. First, it is important to note the overall findings for these three criteria.

In Table 2, findings from all the components in the game-based testing instrument have been compiled into the total calculations for each of the three criteria, with the most important computation being the intra-class correlation coefficient (ICC). When examining the ICC for all three criteria, the outcomes for clarity are the most promising, with an ICC of 0.493. While promising, this finding still does not meet the minimum acceptable ICC level of 0.70 and is not consistent due to diverging confidence intervals for this criterion which echo the high variance result of 0.707 (Tinsley & Weiss, 2000). A lower CI of 0.109 and an upper CI of 0.748 further showcase this high variance amongst the expert ratings. The findings for the latter two factors (ease of use and appropriateness) indicate extremely high variance and low ICC scores. Interestingly, both criteria have negative ICC scores, where ease of use has an ICC of -0.444 and appropriateness an ICC of -0.489. These negative results could be occurring for several reasons, such as high variance, small sample size, negative bias or ICC underestimation (Wu *et al.*, 2012).

Additionally, a closer look into the individual averages amongst the raters demonstrate that two of the experts had very low mean scores throughout the three validity criteria. In contrast, the other experts had means close to the highest rating of five. For example, for the criterion appropriateness, these two experts had average scores of 1.480 and 2.039, and another expert had an average score of 4.88. These results supported the need to conduct follow-up interviews with these two experts. Further, the low ICC scores signal the need to investigate individual scores for the sections within the tool to determine what parts need attention and modification.

Table 2
Combined findings for the criteria clarity, ease of use, and appropriateness including the mean, variance, ICC, and confidence intervals.

	Clarity	Ease of Use	Appropriateness
Mean	3.887	4.025	3.480
Variance	0.707	1.341	2.204
Intraclass Correlation	0.493	-0.444	-0.489
Lower 95% CI	0.109	-1.615	-1.619
Upper 95% CI	0.748	0.300	0.259

Recommendations

Throughout the tool, recommendations are provided to specify how the delivery of the instrument and game design should be fulfilled, such as enlarging, printing, and laminating pictures from a designated list and shuffling the cut-out pictures before testing another participant. As seen in Table 3, this aspect of the modified tool received the highest mean scores for clarity, ease of use, and appropriateness. As a result, much of the recommendations stayed the same.

Table 3
Mean expert scores for the three criteria of face validity relating to the four components assessed within the modified instrument.

	Recommendations	Instructions	Language	Pictures
Clarity	4.41	3.48	3.83	3.36
Ease of Use	3.83	3.75	3.61	3.61
Appropriateness	3.73	3.61	3.44	3.33

Instructions

The second component in the device is the instructions that provide step-by-step guidelines for how to conduct the games, including the associated question prompts and order of operation. Results showcase varied expert ratings due to the means ranging between three and four (Table 3). A three in relation to the five-point scale is a neutral/undecided rating. However, individual experts scores highly fluctuated, such as for the appropriateness of the instructions for game 2b, scores went as low as a one (not clear, not easy to use, and not appropriate) and as high as a five (very clear, very easy to use, and very appropriate).

During an interview with Expert One, it was suggested that the game design and associated instructions for game 2b be revised to include some new elements. In this game, the children are asked if certain pollutants (water pollution, air pollution and ground pollution) can hurt animals, cars, or people. This expert indicated that this game could be taken one step further to include deforestation in the set of impacting factors and the element of “you” and forest in the options for the affected. They reasoned that by including these additional factors there is more substance to analyse, which could then incorporate a discussion about how the children’s answers differ between living versus non-living items getting hurt, and whether they have different emotions and knowledge about things related to the biosphere other than just the three pollutants. For example, a child could believe that deforestation and clear-cutting of trees can hurt the forest, but they may not think water pollution can hurt the forest. This suggestion led to the modification of the tool to include these three new concepts and, therefore, creates an opportunity to highlight further the children’s varying connection to and knowledge of nature.

Language

Another component assessed in relation to the three criteria of face validity is the language used throughout the instrument used to describe the pictures and prompts. While all three criteria received a mean between three and four, not many alterations were suggested by the experts relating to the language used for the pictures and prompts (Table 3). Further, MacKeen and Wright (2020) indicate that the modifications they made, including changes to the language, have enhanced children's understanding of the games testing. However, there were some minor changes made to the phrasing, such as "cut down trees" (MacKeen and Wright, 2020) being altered to "cutting down trees" (latest 2022 version of the tool), since the photo had also been changed from a clear-cut forest to a person standing in a clear cut and cutting down one of the remaining trees. It is important to note that there was a more extensive discussion about the use of language used to create the foundations and construct of the tool, which is later discussed in the results for content validity.

Pictures

The final aspect of the instrument inspected was the photos used to illustrate the concepts. Based on the averages, lack of clarity and appropriateness had the lowest mean scores (Table 3). As later revealed in the results for content validity, certain games were flagged (i.e., game 1b) and then further discussed during the expert interviews. For example, Expert One suggested that in game 1b, the "cleaning up" picture might not have anything to do with environmental relations. This picture was initially modified to a child cleaning the floor with a vacuum indoors (MacKeen and Wright, 2020). The expert proposed that it be related to cleaning up the environment since that is more related to the instrument's construct. Therefore, the photo was changed to children cleaning up garbage outdoors at a beach. Another example in game 1b pertains to the photo of "plastic on the ground," which lacked a clear indication that the plastic is on the ground because it is a zoomed-in picture of a bundle of plastic bottles and cans. As a result, the image now portrays a person with plastic litter on the ground around them.

Content Validity

The results for the scale-level CVI suggest that there are components in the game-based testing instrument that are not adequate for measuring preschool children's affinity with nature. As exhibited in Table 4, the three elements within each game had averages of 0.622 for the instructions, 0.583 for the language, and 0.527 for the pictures. These outcomes are all below the acceptable S-CVI average of 0.80 (Polit *et al.*, 2007). However, these results reiterate issues found in the face validity analysis and indicate that these components have room for improvement that are further identified in the individual CVI scores.

Table 4
Results of the scale-level content validity index averages for the tool components instructions, language, and pictures.

	S-CVI Average
Instructions	0.622
Language	0.583
Pictures	0.527

Instructions

The instructions are included in the instrument to provide researchers with an understanding of how to administer each of the games within the tool. Here is a quote from the instructions for game 1a in the first section of the tool environmental sensitivity:

> *Begin by explaining the exercise to the child:*
> *Example: "In this first game, I will hand you a picture and ask if the thing in the picture can feel an owie or get hurt, and then you will sort them into the yes or no bins (demonstrate while explaining)".*

These explanations are designed to be straightforward for someone else looking to repeat the games. While the instructions received the highest S-CVI average (see Table 4) meaning that the experts felt the instructions were somewhat adequate, the I-CVI provides more information for each of the individual games. As seen in Table 5, the items "before starting the games testing" and "game 1b" meet the acceptable CVI level of 0.80. Game 1a received the next highest score, and games 2a, 2b, 3a, and 3b all received a divisive I-CVI of 0.500, split equally between the experts. These outcomes do not indicate that the instructions are in dire need of further modification. Instead, they suggest they need a second look and a few minor corrections.

Table 5
Outcomes of the interpreted expert ratings and individual content validity index scores for the tool component instructions.

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	I-CVI
Before Starting the Games Testing	0	1	0	0	1	1	0.857
Game 1A	0	1	0	1	1	1	0.667
Game 1B	1	1	0	1	1	1	0.833
Game 2A	0	1	0	0	1	1	0.500
Game 2B	0	1	0	0	1	1	0.500
Game 3A	0	1	0	0	1	1	0.500
Debrief	0	1	0	0	1	1	0.500

Language

The variable "language" encompasses all the language used in the tool, including the recommendations, instructions, words used to describe the pictures and the foundational concepts used to outline the instrument. Our analyses indicate that three out of the six games received a split in opinions from the experts, including games 1b, 2a, and 2b, whereas games 1a, 3a, and 3b got a higher score of 0.667 (Table 5). None of these outcomes meet the acceptable CVI level of 0.80 and suggest more attention is required for games 1b, 2a, and 2b regarding the language choices. However, during interviews with Experts Two and Three it was revealed there was a deeper concern with the language being used to define the construct the instrument intends to measure: bioaffinity.

Table 6
Outcomes of the interpreted expert ratings and individual content validity index scores for the tool component language.

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	I-CVI
Game 1A	0	1	0	1	1	1	0.667
Game 1B	0	1	0	1	1	0	0.500

Game 2A	0	1	0	0	1	1	0.500
Game 2B	0	1	0	0	1	1	0.500
Game 3A	0	1	1	0	1	1	0.667
Game 3B	0	1	1	0	1	1	0.667

One of the key questions asked during the questionnaire is why the experts feel that the games they flag are not relevant for measuring emotional, cognitive, and attitudinal affinity with nature (bioaffinity). Throughout the interview with Expert Three, they continuously came back to whether the games genuinely measure bioaffinity and that the operational definition of bioaffinity is absent. More specifically, for game 1b regarding the photo of water pollution (the photo on the left in Figure 1 below), the expert made the point that "somebody could be very disconnected with nature, but I think they would still find a bird covered in oil sad" and whether it is getting at empathy instead of a child's connection to nature. Another example is given for game 2a, the matching game, and how this showcases the children's knowledge rather than explicitly measuring their bioaffinity. In the same vein, the expert noted that bioaffinity is vague and that these games could measure things that are essential to children's affinity with nature, but it is not clear.

During the final interview session, it was made clear that a thorough review of the domains of bioaffinity was required to ensure that the operationalization of the term includes the items that the tool intends to measure, which are emotional, cognitive, and attitudinal affinity with the biosphere. The final expert profoundly reiterated a lack of congruency between what the title and terminology used in the tool intend to measure and what the games may actually be measuring. For example, game 1b is titled "concern and sensitivity instructions", but the expert indicated that the terms concern and sensitivity are dissimilar and very different from affinity.
Further, they questioned how relevant the terms are for measuring bioaffinity, and it may be that concern and sensitivity foster or lead to affinity or connection, but this assumption should not be inherent and needs additional exploration. In the section for cognitive affinity, the expert highlighted that asking children about ecosystem services is strictly probing their knowledge on the topic as opposed to examining their affinity or connection with nature. It was discussed that assessing their cognition determines the extent of the functions and skills the children possess, such as the ability to memorize a set of words or successfully participate in a sorting task (Sternberg, 1981). As a result of the conversations about the construct with experts three and four, the title of the tool and section headings have been modified (Table 6).

The title and the section headings in the tool specify the concept that is trying to be measured. As such, it is important that the terms used to describe the tool and the games within it reflect the foundational construct the tool aims to measure. The insights gained from the expert interviews led to modifying the original terms to broader concepts that would best embody the intent of the instrument, which is to gauge children's various connections to nature and their environmental knowledge. Consequently, the original title of the tool was revised to "Measuring Environmental Knowledge and Connection to Nature; A Games Testing Tool for Preschoolers (3-5-year-olds)" (see Table 6).

Table 6
Construct related terminology modifications.

Element in the Tool	**Original**	**Modification**
Title of the tool	Modified Research Instrument (Games Testing for Emotional, Cognitive and Attitudinal Affinity with the Biosphere, Giusti *et al.*, 2014)	Measuring Environmental Knowledge and Connection to Nature; A Games Testing Tool for Preschoolers (3-5-year-olds)
Title of Game 1A	Emphatic Behavior Instructions	Environmental Sensitivity

Title of Game 1B	Concern & Sensitivity Instructions	
Title of Game 2A	Provision of Ecosystem Services Instructions	Environmental Awareness
Title of Game 2B	Pollution Awareness Instructions	
Title of Game 3A	Favorite Environmental Quality Instructions	Environmental Preferences
Title of Game 3B	Disfavored Environmental Quality Instructions	

Additionally, the section headings and titles of the games have been changed from the MacKeen and Wright (2020) version of the tool. Initially there were three sections as identified in the original title: emotional affinity, cognitive affinity, and attitudinal affinity. As seen in Table 6, each game had their own names and were placed within each of the three segments. These findings highlighted the confusion between headings and the names of the games, such as game 1a, emphatic behaviour instructions, being placed in what was originally the emotional affinity sector. The term emphatic is defined in multiple variations, for example "uttered with or marked by emphasis", "tending to express oneself in forceful speech or to take decisive action ", and "attracting special attention" (Merriam-Webster, 2022c). This confusion amongst terminology is evident for other language that was used to build the instrument. The following discussion will further explore the terminology and the foundational constructs of the tool.

Pictures

The final component pictures include responses about the images used to depict various items used to measure the children's connection to nature, such as a photo of water pollution. This category received the lowest overall S-CVI average of 0.527 (see Table 3). Taking a closer look, it is apparent that game 2a has the lowest I-CVI of 0.333 (Table 7). Game 2a is a matching game where children are asked to match items to where they come from, hoping they will pair the items with the environmental option (i.e., matching blueberries with a garden). With four out of the six experts rating this a three or below on the Likert scale, it is clear that further modification is necessary for the photos within this game. Games 1a, 1b, and 2b had split scores from the experts of 0.500 and games 3a, and 3b received the highest I-CVI for this criteria (see Table 7).

Table 7
Outcomes of the interpreted expert ratings and individual content validity index scores for the tool component pictures.

	Expert 1	**Expert 2**	**Expert 3**	**Expert 4**	**Expert 5**	**Expert 6**	**I-CVI**
Game 1A	0	1	0	1	1	0	0.500
Game 1B	0	1	0	0	1	1	0.500
Game 2A	0	1	0	0	1	0	0.333
Game 2B	0	1	0	0	1	1	0.500
Game 3A	0	1	1	0	1	1	0.667
Game 3B	0	1	1	0	1	1	0.667

These three criteria echo the findings from face validity and suggest that the contents of the instrument are not adequate for measuring affinity with nature, which was further supported during the interviews with the three experts. Expert Three made some crucial points about how it could be difficult for the children to understand what is happening in such complex photos relating to environmental degradation. For example, as seen in Figure 1, the picture on the left with a bird being pulled from an oil spill portrays water pollution. However, it is understandable

that this picture may not be developmentally appropriate for such a young age group, primarily because of the assumption that the children can understand the antecedence leading to the result of a bird being covered in oil. More specifically, it is unwarranted to assume all of the 3-5-year-old children can connect this to a human-caused oil spill. This led to the revision of the picture to include a person physically putting litter into a waterway to make the photo more digestible for the target audience (see Figure 1).

Another concern for this expert is the picture initially used to depict air pollution. This picture had been modified in 2019 to a photo of a person wearing a mask with smoke in the background, but the photo has not aged well with the ongoing COVID-19 pandemic. It was discussed that masks would now be commonly associated with the virus rather than protection from air pollution or any form of smoke. As a result, the photo has been exchanged for a photo of cars driving along a roadway with a cloud of smog in the air, in hopes that this illustration could be better associated with how automobiles contribute to air pollution.

Figure 1. Original (MacKeen and Wright, 2020) and modified (2022) pictures that illustrate water pollution.

Overall, the face and content validity findings and support from the interviews suggested there were cosmetic and foundational issues throughout the instrument. The following section will further explore the theoretical concepts and construct of the device and the problem with disguising and manipulating definitions to better suit individual research**.**

Discussion

Within any individual scholarly discipline, a researcher can expect to find concepts, methodologies, theories, topics, and terms engineered explicitly for use within that specific school of thought (Stock & Burton, 2011). However, these characteristics can become blurred, masked, and absent when they are not fully understood or explained (Stock & Burton, 2011). In the case of developing psychological instruments, strong definitions and explanations of concepts are critical to establishing a solid construct (Stone-Romero *et al.*, 2009). The foundations of the instrument we have been using and which was subject to validity testing in this study were first conceptualized in a 2012 manuscript titled: Reconnecting to the Biosphere; Children's Socio-ecological Emotions for Nature (Giusti, 2012). Inner workings of the framework used to build the foundation consisted of exploring how different socio-ecological environments influence the development of environmental consciousness in children. Environmental consciousness in the context of this study was defined as "a set of psychological traits held by an individual which specifically represent the individual emotional connection with the biosphere" (Giusti, pg. 12, 2012). It was said that environmental consciousness consisted of four characteristics, including environmental sensitivity, awareness, attitude, and ethics.

Environmental sensitivity is described as a "conjunction of empathy and concern, as caring for a person implies also being concerned about his or her health" (Giusti, pg. 24, 2012). Environmental awareness is said to include a cognitive, knowledge-based component and an affective, perception-based component that is not limited to the impacts of human behaviour on the environment but also knowledge about essential ecosystem services and nature (Giusti, 2012). Environmental attitude in this manuscript is defined as "a durable positive or negative feeling towards the biosphere, regardless of the deriving behaviour" (Giusti, pg. 25, 2012). Finally, environmental ethics is described as personal values and beliefs about the environment that influence environmental behaviours (Giusti, 2012). Results from Giusti's study indicated that children with higher exposure to wild and rural environments also have higher empathy and concern for nature and led to the creation of the instrument (Games Testing for Emotional, Cognitive and Attitudinal Affinity with the Biosphere, Giusti *et al.*, 2014) to further explore these findings.

However, during the composition of the Giusti *et al.* (2014) tool, these fundamental terms and ideas were restated and potentially lost in translation. As previously noted, the tool is divided into three sections, including emotional affinity, cognitive affinity, and attitudinal affinity. Emotional affinity with the biosphere is assembled to quantify the children's emotional perspective-taking and empathetic concern for nature (Giusti *et al.*, 2014). Cognitive affinity with the biosphere targets measuring the children's basic awareness of ecological resources, and attitudinal affinity explores their appreciation of nature and environmental awareness (Giusti *et al.*, 2014). It is clear that there is a mixing and perplexing misuse of specific terms, such as using environmental awareness with attitudinal affinity and even the simplicity of using the term affinity instead of connection. These terms have different meanings and definitions, and while researchers can manipulate and define them for their studies, the question remains: should we? This highlights how elements of theories become misconstrued and replaced by surrogates. This can in turn lead to weak construct validity in the foundation of the research due to imprecise theoretical components (Stone-Romero *et al.*, 2009).

The method of circular restatement is known to avoid theories and concepts by restating phenomena in different words, causing tautology (Gigerenzer, 2010). When researchers restate and provide one-word explanations of the construct, the foundation of the device and its intentions come into question (Gigerenzer, 2010). Further, using elaborate and decorative terms over simple ones impacts how the research connects to the broader community for the targeted discipline, and other validity tests (i.e., face and content validity) rely on a sound construct to properly assess the tool's functions (Cronbach & Meehl, 1955). Therefore, comprehensive theoretical components used to establish psychological behavioural assessments are necessary for evaluating the device's psychometrics (i.e., validity and reliability).

Giusti *et al.*, (2014) wanted to "analyze the extinction of nature experiences in the surrounding of urban preschools in Stockholm and relates it to the degree of affinity with the biosphere that 5-year-old children have developed" (pg. 18). As such, the premise of the Giusti *et al.* (2014) tool is to measure children's affinity with nature, but the question remains: what is affinity? Affinity can be broadly defined as "a spontaneous or natural liking or sympathy for someone or something" and "a similarity of characteristics suggesting a relationship" (Oxford Languages, 2022a). This is a loose term that can be spontaneous and merely suggest a relationship, implying that it may not identify the degree of affinity because the meaning of the word is not static. Affinity is also not a very common term in environmental psychology. The Giusti *et al.* (2014) tool is inspired by other assessments, including the Connectedness to Nature Scale (Mayer & Frantz, 2004), the Nature Relatedness Scale (Nisbet *et al.,* 2009), and the Connection to Nature index (Cheng & Monroe, 2012). However, the terminology has strayed from common terms used in this line of research, such as connection and relations with nature. In comparison to affinity, a connection is defined as "a relationship in which a person, thing, or idea is linked or associated with something else" (Oxford Languages, 2022b). While affinity suggests a relationship, connection links a person with something, indicating a direct relationship. These differences may cause the objective and intent of the construct to get blurred and restated to something not originally envisioned.

After examining the disconnect between the 2012 and 2014 Giusti texts and conferring with the experts, it is important to peel back the layers of the instrument and redefine the foundational construct and concepts. The premise for all versions of the tool seek to explore children's connection to nature and how their exposure to outdoor environments and environmental knowledge affect such a relationship. This resulted in the adoption of the term

connection in place of affinity. However, many of the games are also a measure of the children's knowledge because, without comprehending the material, there would be no way to analyze their relationship with nature successfully. Thus, the core construct of the tool is now measuring children's connection to and knowledge of nature and its processes, which leads to the title of the newly improved tool being: Measuring Environmental Knowledge and Connection to Nature; A Games Testing Tool for Preschoolers (3-5-year-olds). Now that the central theory of the tool has changed, the subsections need to adhere to this change.

Another element at play in these revisions was the Giusti (2012) manuscript that outlines the four characteristics of environmental consciousness as mentioned above. The basis of these features (sensitivity, awareness, attitudes, and ethics) are topics that are frequently investigated in psychological research. Children's sensitivity is explored in correlation to facial expressions (Gao & Maurer, 2009), musical styles (Gardner, 1973), and word use (Markman & Hutchinson, 1984) as few examples. Their awareness is measured relating to matters such as awareness of inconsistent information (Markman, 1979), fatal illnesses (Waechter, 1971), and internet safety (Ktoridou *et al.*, 2012). Many studies measure various aspects of children's attitudes, including their attitudes towards reading (McKenna *et al.*, 1995), pets (Kidd & Kidd, 1985), peers (Coplan *et al.*, 2007), healthy eating (Bebetsos *et al.*, 2015) and much more. Finally, children's ethics or moral development is also a topic of interest in various studies (Berkowitz & Grych, 1998; Eisenberg & Valiente, 1995; Kurdek, 1978; Turiel, 2015). However, in 2014 when Giusti *et al.* established the tool, these topics were disregarded and replaced with emotional affinity, cognitive affinity, and attitudinal affinity with nature. While on the surface these terms seem to encompass some of the other features, such as sensitivity being associated with emotions, not all of these categories and expressions are commonly explored and used in environmental psychology.

Additionally, the games within these categories claim to measure different things than what is suggested by the overarching section. For example, in game 1b, the concern and sensitivity instructions include two terms where concern can be defined as "to relate to" or "to have an influence on", and sensitive as being "delicately aware of the attitudes and feelings of others" (Merriam-Webster, 2022a, 2022b). While these two concepts are explored in environmental psychology research, they are most commonly addressed separately; it is far more common to find studies that utilize environmental sensitivity in conjunction with research that includes children (Cheng & Wu, 2015; Chu *et al.*, 2007; Erdogan & Marcinkowski, 2015; Lionetti *et al.*, 2019; Nocentini *et al.*, 2018). Another example is game 3a, labeled as favourite environmental quality instructions, in section three under attitudinal affinity. The name of the game was previously disconnected from the concept (attitude) that is to be measured for this segment of the instrument. The word favourite may be a sentiment that drives the child's attitude towards the environment, but the word quality distracts from the principal idea. As a result of these disputes, the titles for each of the six games have been removed, and the overarching sections have been changed to reflect the initial concepts used to inspire the makeup of the instrument (Table 6).

Environmental sensitivity, awareness, and preferences are fundamental beliefs that foster one's connection to nature and environmental knowledge. It is important to note that environmental ethics has been omitted from the tool due to the targeted audience and their limited developmental capacity for moral sentiments. It is essential that as psychological instruments emerge in specialized disciplines, the theories and terms used to describe that concept are clearly defined and are suitable for the target audience and the greater field of research. By modifying the foundational aspects, it will allow the device to become more widespread in its universal applicability and placement in environmental psychology. Finally, now that the construct of the tool has been revealed, the instrument can undergo further psychometric evaluation and produce trustworthy outcomes.

Conclusions and Future Implications

Establishing the validity of the modified instrument for measuring preschool children's connection to nature and environmental knowledge is a critical step towards a psychometrically sound and trustworthy device. This study sought to explore the face and content validity of the revised (2020) version of this device and determine whether the individual and overall components aid in quantifying the construct of the tool. The results revealed low face and content scores, which led to three follow-up expert interviews and further modification of the instrument. Interviews

with the experts exposed a disconnect between what the tool intends to measure and the characteristics used to outline the features relied upon to measure the core concept(s). Therefore, the foundations and underlying inspirations of the instrument were investigated and led to a discussion about the importance of utilizing well-defined terminology to frame a psychological device. As such, the construct of the tool has been revised to measure children's connection to nature and environmental knowledge, with the internal sections reflecting the new title by changing them to environmental sensitivity, awareness and preferences. Alterations to the pictures, language and instructions also took place to enhance the clarity, ease of use, appropriateness and relevancy for measuring children's connection to nature and environmental knowledge.

However, it is important to note that the face validity outcomes are no longer applicable due to the tool undergoing further revisions. This limitation implies that face validity would need to be re-evaluated for the new 2022 version of the modified instrument, and/or other facets of validity testing (for example, criterion validity) need to take place before the tool is used in a larger format study. Despite this, we suggest that future studies that seek to use this instrument use the newly modified 2022 version as it is now the most relevant for measuring preschoolers connection to nature and environmental knowledge (please note that a full copy of the instrument is available by contacting the authors or by visiting this link).

Moreover, the hope for this instrument is for more research to explore how various forms of nature exposure impact preschoolers' CTN and environmental knowledge, such as a comparison of CTN between children who attend nature-based (i.e., Reggio-Emilia inspired) versus non-natured based schools. More research concerning the biological and developmental growth of 3-5-year-olds and cultural and geographical influences is imperative as the tool is used in various locations worldwide. Finally, this modified instrument creates a lasting impact in the field of environmental psychology as it should be considered a living tool that is manipulated to suit different geographic, cultural, and young developmental stages. By establishing such a tool, scholars can use this chain of research as a guide for how to develop, modify, and psychometrically evaluate game-based testing instruments.

Declaration of Conflicting Interests

It is declared that there are no conflicts of interest relating to the research, authorship, and publication of this article.

Funding

The graduate student Jessica MacKeen discloses that her research has been partially funded by the Faculty of Graduate Studies at Dalhousie University.

References

Bajpai, S., Bajpai, R., & Chaturvedi, H. (2015). Evaluation of Inter-Rater Agreement and Inter-Rater Reliability for Observational Data: An Overview of Concepts and Methods. *Journal of the Indian Academy of Applied Psychology*, *41*, 20–27.

Barrable, A., & Booth, D. (2020). Increasing Nature Connection in Children: A Mini Review of Interventions. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00492

Bartko, J. J. (1966). The Intraclass Correlation Coefficient as a Measure of Reliability. *Psychological Reports*, *19*(1), 3–11. https://doi.org/10.2466/pr0.1966.19.1.3

Bebetsos, E., Zorzou, A., Bebetsos, G., Kosta, G., & Georgios, K. (2015). Children's self-efficacy and attitudes towards healthy eating. An application of the Theory of Planned Behavior. *International Journal of Sports and Physical Education*, *1*, 1–8.

Berkowitz, M. W., & Grych, J. H. (1998). Fostering Goodness: Teaching parents to facilitate children's moral development. *Journal of Moral Education*, *27*(3), 371–391. https://doi.org/10.1080/0305724980270307

Braus, J., & Milligan-Toffler, S. (2018). The Children and Nature Connection: Why It Matters. *Ecopsychology*, *10*(4), 193–194. https://doi.org/10.1089/eco.2018.0072

Buyukkarci, K. (2009). A critical analysis of task-based learning. *Kastamonu Journal of Education*, *17*, 313–320.

Chawla, L. (2015). Benefits of Nature Contact for Children. *Journal of Planning Literature*, *30*(4), 433–452. https://doi.org/10.1177/0885412215595441

Cheng, J. C.-H., & Monroe, M. C. (2012). Connection to Nature: Children's Affective Attitude Toward Nature. *Environment and Behavior*, *44*(1), 31–49. https://doi.org/10.1177/0013916510385082

Cheng, T.-M., & Wu, H. C. (2015). How do environmental knowledge, environmental sensitivity, and place attachment affect environmentally responsible behavior? An integrated approach for sustainable island tourism. *Journal of Sustainable Tourism*, *23*(4), 557–576. https://doi.org/10.1080/09669582.2014.965177

Chipeniuk, R. (1995). Childhood Foraging as a Means of Acquiring Competent Human Cognition about Biodiversity. *Environment and Behavior*, *27*(4), 490–512. https://doi.org/10.1177/0013916595274003

Chu, H., Lee, E. A., Ryung Ko, H., Hee Shin, D., Nam Lee, M., Mee Min, B., & Hee Kang, K. (2007). Korean Year 3 Children's Environmental Literacy: A prerequisite for a Korean environmental education curriculum. *International Journal of Science Education*, *29*(6), 731–746. https://doi.org/10.1080/09500690600823532

Clayton, S., & Opotow, S. (2003). *Identity and the Natural Environment: The Psychological Significance of Nature*. MIT Press.

Cohen, L., Manion, L., & Morrison, K. (2002). *Research Methods in Education*. Taylor & Francis Group. http://ebookcentral.proquest.com/lib/dal/detail.action?docID=170477

Connell, J., Carlton, J., Grundy, A., Taylor Buck, E., Keetharuth, A. D., Ricketts, T., Barkham, M., Robotham, D., Rose, D., & Brazier, J. (2018). The importance of content and face validity in instrument development: Lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL). *Quality of Life Research*, *27*(7), 1893–1902. https://doi.org/10.1007/s11136-018-1847-y

Coplan, R. J., Girardi, A., Findlay, L. C., & Frohlick, S. L. (2007). Understanding Solitude: Young Children's Attitudes and Responses Toward Hypothetical Socially Withdrawn Peers. *Social Development*, *16*(3), 390–409. https://doi.org/10.1111/j.1467-9507.2007.00390.x

Cronbach, L. J., & Meehl, P. E. (1955). CONSTRUCT VALIDITY IN PSYCHOLOGICAL TESTS. *Psychological Bulletin*, *52*, 281–302.

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale Development: Theory and Applications*. SAGE Publications.

Dijkstra, E. M., & Goedhart, M. J. (2012). Development and validation of the ACSI: Measuring students' science attitudes, pro-environmental behaviour, climate change attitudes and knowledge. *Environmental Education Research*, *18*(6), 733–749. https://doi.org/10.1080/13504622.2012.662213

Dresner, M., Handelman, C., Braun, S., & Rollwagen-Bollens, G. (2015). Environmental identity, pro-environmental behaviors, and civic engagement of volunteer stewards in Portland area parks. *Environmental Education Research*, *21*(7), 991–1010. https://doi.org/10.1080/13504622.2014.964188

Eisenberg, N., & Valiente, C. (1995). Parenting and Children's Prosocial and Moral Development. In B. Tinsley & N. Lees (Eds.), *Handbook of Parenting* (Vol. 4, pp. 148–179).

Erdogan, M., & Marcinkowski, T. (2015). Development and Validation of Children's Environmental Affect (Attitude, Sensitivity and Willingness to Take Action) Scale. *EURASIA Journal of Mathematics, Science and Technology Education*, *11*(3). https://doi.org/10.12973/eurasia.2015.1347a

Ewert, A., Place, G., & Sibthorp, J. (2005). Early-Life Outdoor Experiences and an Individual's Environmental Attitudes. *Leisure Sciences*, *27*(3), 225–239. https://doi.org/10.1080/01490400590930853

Gao, X., & Maurer, D. (2009). Influence of intensity on children's sensitivity to happy, sad, and fearful facial expressions. *Journal of Experimental Child Psychology*, *102*(4), 503–521. https://doi.org/10.1016/j.jecp.2008.11.002

Gardner, H. (1973). CHILDREN'S SENSITIVITY TO MUSICAL STYLES. *Merrill-Palmer Quarterly of Behavior and Development*, *19*(1), 67–77.

Gigerenzer, G. (2010). Personal Reflections on Theory and Psychology. *Theory & Psychology*, *20*(6), 733–743. https://doi.org/10.1177/0959354310378184

Giusti, M. (2012). *Reconnecting to the Biosphere: Children's Socio-Ecological Emotions for Nature (Published Master's Thesis).*

Giusti, M., Barthel, S., & Marcus, L. (2014). Nature Routines and Affinity with the Biosphere: A Case Study of Preschool Children in Stockholm. *Children, Youth and Environments*, *24*(3), 16–42.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*(3), 238–247. https://doi.org/10.1037/1040-3590.7.3.238

Holden, R. R. (2010). Face Validity. In *The Corsini Encyclopedia of Psychology* (pp. 1–2). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470479216.corpsy0341

Kahn, P., & Kellert, S. R. (2002). *Children and Nature: Psychological, Sociocultural, and Evolutionary Investigations*. MIT Press.

Kallogjeri, D., Francis, D. O., & Piccirillo, J. F. (2019). Use of Confidence Intervals to Improve the Quality of Research Results Reporting and Interpretation. *JAMA Otolaryngology–Head & Neck Surgery*, *145*(6), 491–493. https://doi.org/10.1001/jamaoto.2019.0793

Kellert, S. R. (2012). *Building for Life: Designing and Understanding the Human-Nature Connection*. Island Press.

Kidd, A. H., & Kidd, R. M. (1985). Children's Attitudes toward Their Pets. *Psychological Reports*, *57*(1), 15–31. https://doi.org/10.2466/pr0.1985.57.1.15

Krause, M. S. (2012). Measurement Validity is Fundamentally a Matter of Definition, Not Correlation. *Review of General Psychology*, *16*(4), 391–400. https://doi.org/10.1037/a0027701

Ktoridou, D., Eteokleous, N., & Zahariadou, A. (2012). Exploring parents' and children's awareness on internet threats in relation to internet safety. *Campus-Wide Information Systems*, *29*(3), 133–143. https://doi.org/10.1108/10650741211243157

Kurdek, L. A. (1978). PERSPECTIVE TAKING AS THE COGNITIVE BASIS OF CHILDREN'S MORAL DEVELOPMENT: A REVIEW OF THE LITERATURE. *Merrill-Palmer Quarterly of Behavior and Development*, *24*(1), 3–28.

Larson, L. R., Green, G. T., & Castleberry, S. B. (2011). Construction and Validation of an Instrument to Measure Environmental Orientations in a Diverse Group of Children. *Environment and Behavior*, *43*(1), 72–89. https://doi.org/10.1177/0013916509345212

Limpo, T., Filipe, M., Magalhães, S., Cordeiro, C., Veloso, A., Castro, S. L., & Graham, S. (2020). Development and validation of instruments to measure Portuguese third graders' reasons to write and self-efficacy. *Reading and Writing*, *33*(9), 2173–2204. https://doi.org/10.1007/s11145-020-10039-z

Lionetti, F., Aron, E. N., Aron, A., Klein, D. N., & Pluess, M. (2019). Observer-Rated Environmental Sensitivity Moderates Children's Response to Parenting Quality in Early Childhood. *Developmental Psychology*, *55*(11), 2389–2402. https://doi.org/10.1037/dev0000795

Littlewood, W. (2004). The task-based approach: Some questions and suggestions. *ELT Journal*, *58*(4), 319–326. https://doi.org/10.1093/elt/58.4.319

MacKeen, J., & Wright, T. (2020). Refining a Games Testing Tool for Various Cultural, Social, and Geographic Situations to Evaluate Preschool Children's Bioaffinity. *The International Journal of Early Childhood Environmental Education*, *8*(1), 17.

Malone, K., & Tranter, P. (2003). "Children's Environmental Learning and the Use, Design and Management of Schoolgrounds. *Children, Youth and Environments*, *13*(2), 87–137.

Markman, E. M. (1979). Realizing That You Don't Understand: Elementary School Children's Awareness of Inconsistencies. *Child Development*, *50*(3), 643–655. https://doi.org/10.2307/1128929

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, *16*(1), 1–27. https://doi.org/10.1016/0010-0285(84)90002-1

Mayer, F. S., & Frantz, C. M. (2004). The connectedness to nature scale: A measure of individuals' feeling in community with nature. *Journal of Environmental Psychology*, *24*(4), 503–515. https://doi.org/10.1016/j.jenvp.2004.10.001

McCurdy, L. E., Winterbottom, K. E., Mehta, S. S., & Roberts, J. R. (2010). Using Nature and Outdoor Activity to Improve Children's Health. *Current Problems in Pediatric and Adolescent Health Care*, *40*(5), 102–117. https://doi.org/10.1016/j.cppeds.2010.02.003

McKenna, M. C., Kear, D. J., & Ellsworth, R. A. (1995). Children's Attitudes toward Reading: A National Survey. *Reading Research Quarterly*, *30*(4), 934–956. https://doi.org/10.2307/748205

Merriam-Webster. (2022a). *Definition of CONCERN*. https://www.merriam-webster.com/dictionary/concern

Merriam-Webster. (2022b). *Definition of SENSITIVE*. https://www.merriam-webster.com/dictionary/sensitive

Merriam-Webster. (2022c, February 14). *Emphatic Definition & Meaning—Merriam-Webster*. https://www.merriam-webster.com/dictionary/emphatic

Mygind, L., Kjeldsted, E., Hartmeyer, R., Mygind, E., Bølling, M., & Bentsen, P. (2019). Mental, physical and social health benefits of immersive nature-experience for children and adolescents: A systematic review and quality assessment of the evidence. *Health & Place*, *58*, 102136. https://doi.org/10.1016/j.healthplace.2019.05.014

Nisbet, E. K., Zelenski, J. M., & Murphy, S. A. (2009). The Nature Relatedness Scale: Linking Individuals' Connection With Nature to Environmental Concern and Behavior. *Environment and Behavior*, *41*(5), 715–740. https://doi.org/10.1177/0013916508318748

Njelesani, J., Davis, J. A., & Pontes, T. (2020). Occupational Repertoire Development Measure – Parent (ORDM-P): Face validity, comprehensiveness, and internal consistency. *British Journal of Occupational Therapy*, *83*(5), 326–333. https://doi.org/10.1177/0308022619885247

Nocentini, A., Menesini, E., & Pluess, M. (2018). The Personality Trait of Environmental Sensitivity Predicts Children's Positive Response to School-Based Antibullying Intervention. *Clinical Psychological Science*, *6*(6), 848–859. https://doi.org/10.1177/2167702618782194

Omidvar, N. (2018). *Investigating the Effect of Nature-Related Routines on Preschool Children's Affinity to Nature at Halifax Children's Centres (Published Master's Thesis).* Dalhousie Univeristy.

Omidvar, N., Wright, T., Beazley, K., & Seguin, D. (2019). Examining Children's Indoor and Outdoor Nature Exposures and Nature-related Pedagogic Approaches of Teachers at Two Reggio-Emilia Preschools in Halifax, Canada. *Journal of Education for Sustainable Development*, *13*(2), 215–241. https://doi.org/10.1177/0973408219872066

Oxford Languages. (2022a). *define affinity—Google Search*. https://www.google.com/search?q=define+affinity&oq=define+affinity&aqs=chrome..69i57j0i512l4j0i20i263i512j0i512l4.3106j1j7&sourceid=chrome&ie=UTF-8

Oxford Languages. (2022b). *define connection—Google Search*. https://www.google.com/search?q=define+connection&oq=define+connection&aqs=chrome..69i57j0i512l9.3570j1j7&sourceid=chrome&ie=UTF-8%27

Piredda, M., Ghezzi, V., Fenizia, E., Marchetti, A., Petitti, T., De Marinis, M. G., & Sili, A. (2017). Development and psychometric testing of a new instrument to measure the caring behaviour of nurses in Italian acute care settings. *Journal of Advanced Nursing*, *73*(12), 3178–3188. https://doi.org/10.1111/jan.13384

Polit, D. F., & Beck, C. T. (2009). *Essentials of Nursing Research: Appraising Evidence for Nursing Practice*. Lippincott Williams & Wilkins.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, *30*(4), 459–467. https://doi.org/10.1002/nur.20199

Rickinson, M. (2001). Learners and Learning in Environmental Education: A critical review of the evidence. *Environmental Education Research*, *7*(3), 207–320. https://doi.org/10.1080/13504620120065230

Roberts, P., Priest, H., & Traynor, M. (2006). Reliability and validity in research. *Nursing Standard (through 2013)*, *20*(44), 41–45.

Schultz, P. W. (2002). Inclusion with Nature: The Psychology Of Human-Nature Relations. In P. Schmuck & W. P. Schultz (Eds.), *Psychology of Sustainable Development* (pp. 61–78). Springer US. https://doi.org/10.1007/978-1-4615-0995-0_4

Shi, J., Mo, X., & Sun, Z. (2012). [Content validity index in scale development]. *Zhong nan da xue xue bao Yi xue ban = Journal of Central South University. Medical sciences*, *37*(2), 152–155. https://doi.org/10.3969/j.issn.1672-7347.2012.02.007

Silva, W. A. D., de Sampaio Brito, T. R., & Pereira, C. R. (2020). COVID-19 anxiety scale (CAS): Development and psychometric properties. *Current Psychology*. https://doi.org/10.1007/s12144-020-01195-0

Singh, R., & Awasthi, S. (2020). Updated Comparative Analysis on Video Conferencing Platforms – Zoom, Google Meet, Microsoft Teams, WebEx Teams and GoToMeetings. EasyChair Preprint, 4026. *EasyChair Preprint, 4026.*

Sireci, S. G. (1998). The Construct of Content Validity. *Social Indicators Research*, *45*, 83–117.

Slater, P., McCormack, B., & Bunting, B. (2009). The Development and Pilot Testing of an Instrument to Measure Nurses' Working Environment: The Nursing Context Index. *Worldviews on Evidence-Based Nursing*, *6*(3), 173–182. https://doi.org/10.1111/j.1741-6787.2009.00159.x

Stock, P., & Burton, R. J. F. (2011). Defining Terms for Integrated (Multi-Inter-Trans-Disciplinary) Sustainability Research. *Sustainability*, *3*(8), 1090–1113. https://doi.org/10.3390/su3081090

Stone-Romero, E. F., Alvarez, K., & Thompson, L. F. (2009). The construct validity of conceptual and operational definitions of contextual performance and related constructs. *Human Resource Management Review*, *19*(2), 104–116. https://doi.org/10.1016/j.hrmr.2008.10.003

Tam, K.-P. (2013). Concepts and measures related to connection to nature: Similarities and differences. *Journal of Environmental Psychology*, *34*, 64–78. https://doi.org/10.1016/j.jenvp.2013.01.004

Tinsley, H. E. A., & Weiss, D. J. (2000). 4—Interrater Reliability and Agreement. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (pp. 95–124). Academic Press. https://doi.org/10.1016/B978-012691360-6/50005-7

Turiel, E. (2015). Moral development. In *Handbook of child psychology and developmental science: Theory and method, Vol. 1, 7th ed* (pp. 484–522). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118963418.childpsy113

Waechter, E. H. (1971). Children's Awareness of Fatal Illness. *The American Journal of Nursing*, *71*(6), 1168–1172. https://doi.org/10.2307/3421945

Wu, S., Crespi, C. M., & Wong, W. K. (2012). Comparison of methods for estimating the intraclass correlation coefficient for binary responses in cancer prevention cluster randomized trials. *Contemporary Clinical Trials*, *33*(5), 869–880. https://doi.org/10.1016/j.cct.2012.05.004

Zhang, W., Goodale, E., & Chen, J. (2014). How contact with nature affects children's biophilia, biophobia and conservation attitude in China. *Biological Conservation*, *177*, 109–116. https://doi.org/10.1016/j.biocon.2014.06.011

Jessica MacKeen is a Master of Environmental Studies Graduate in the School for Resource and Environmental Studies at Dalhousie University, Nova Scotia, Canada. She may be reached at jmackeen@dal.ca.

Tarah Wright is a Professor in the Department of Earth and Environmental Sciences at Dalhousie University, Nova Scotia, Canada. She can be reached at tarah.wright@dal.ca.

Daniel Séguin is a Professor of Psychology at Mount Saint Vincent University and can be reached at daniel.seguin@msvu.ca.

Heather Cray is an Instructor in the School for Resource and Environmental Studies at Dalhousie University, Nova Scotia, Canada. She can be reached at hcray@dal.ca.