





## Critical Review

# Conceptualization of the term “ecological validity” in neuropsychological research on executive function assessment: a systematic review and call to action

Yana Suchy<sup>1</sup> , Libby A. DesRuisseaux<sup>1</sup> , Michelle Gereau Mora<sup>1</sup>, Stacey Lipio Brothers<sup>1</sup> and Madison A. Niermeyer<sup>2</sup>

<sup>1</sup>Department of Psychology, University of Utah, Salt Lake City, UT, USA and <sup>2</sup>Department of Physical Medicine and Rehabilitation, University of Utah, Salt Lake City, UT, USA

### Abstract

**Objective:** “Ecological validity” (EV) is classically defined as test’s ability to predict real-world functioning, either alone or together with test’s similarity to real-world tasks. In neuropsychological literature on assessment of executive functions (EF), EV is conceptualized inconsistently, leading to misconceptions about the utility of tests. The goal of this systematic review was to examine how EV is conceptualized in studies of EF tests described as ecologically valid. **Method:** MEDLINE and PsychINFO Databases were searched. PRISMA guidelines were observed. After applying inclusion and exclusion criteria, this search yielded 90 articles. Deductive content analysis was employed to determine how the term EV was used. **Results:** About 1/3 of the studies conceptualized EV as the test’s ability to predict functional outcomes, 1/3 as both the ability to predict functional outcome and similarity to real-world tasks, and 1/3 were either unclear about the meaning of the term or relied on notions unrelated to classical definitions (e.g., similarity to real-world tasks alone, association with other tests, or the ability to discriminate between populations). **Conclusions:** Conceptualizations of the term EV in literature on EF assessment vary grossly, subsuming the notions of criterion, construct, and face validity, as well as sensitivity/specificity. Such inconsistency makes it difficult to interpret clinical utility of tests that are described as ecologically valid. We call on the field to require that, at minimum, the term EV be clearly defined in all publications, or replaced with more concrete terminology (e.g., criterion validity).

**Keywords:** executive functions; instrumental activities of daily living; face validity; predictive validity; criterion validity

(Received 10 May 2023; final revision 3 October 2023; accepted 5 October 2023; First Published online 22 January 2024)

The term “ecological validity” (EV) has been defined variably across years and disciplines. It was originally coined in the 1940s by Egon Brunswik, pertaining to the degree to which a percept provides information about the actual properties of perceived stimulus (Brunswik, 1956). In the 1960s and 1970s, experimental psychologists began to use EV to reflect the degree to which an experimental manipulation paralleled real-world cause-and-effect relationships (Anisfeld, 1968; Dudycha et al., 1973; Jennings & Keefer, 1969); and by the 1980s, clinical and developmental psychologists began to apply EV to intelligence testing, questioning whether IQ scores alone could explain real-world functioning (Gaylord-Ross, 1979; Latham, 1978; Wiedl & Herrig, 1978). On the heels of these developments, the emerging field of clinical neuropsychology began to question its own assessment methods (Newcombe, 1987), leading to a flurry of ecologically-themed publications in the early 1990s (Farmer & Eakman, 1995; Gass et al., 1990; Johnson, 1994; Wilson, 1993), and culminating with the publication of a prominent edited textbook fully devoted to EV of neuropsychological assessment (Sbordone & Long, 1996). As

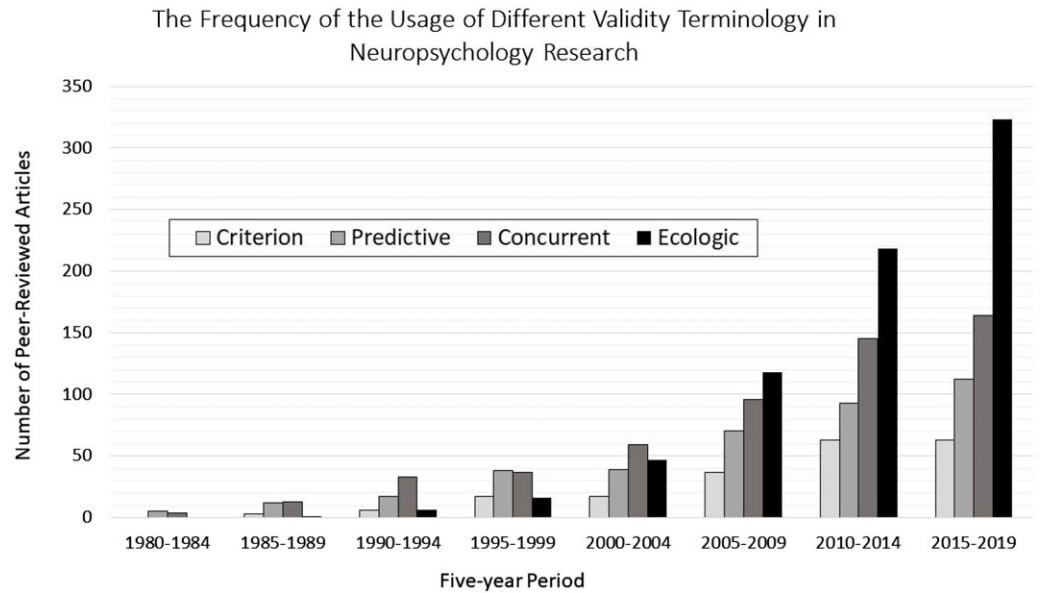
seen in Figure 1, following the publication of Sbordone’s and Long’s (1996) book, the term *EV* took a firm hold in the neuropsychological literature, and has since eclipsed the usage of other well-established validity terms, including predictive, concurrent, or criterion validity.

Given the proliferation of literature that examines, criticizes, or otherwise discusses EV of neuropsychological instruments, one would expect the term to be well understood and used consistently across studies. However, even a casual perusal of the literature reveals considerable inconsistencies. On the one hand, in their 1996 textbook, Sbordone (1996, p. 16) defined EV as “the functional and predictive relationship between the patient’s performance . . . and the patient’s behavior in a variety of real-world settings” (i.e., the ability to predict real-world outcomes), which Long (1996) echoed. On the other hand, in the same text, Franzen & Wilhelm (1996) proposed a two-pronged conceptualization of EV, stating that ecological validation involves “investigations of *both* verisimilitude *and* veridicality” (p. 96, italics added), wherein “verisimilitude” refers to “the similarity of the data

**Corresponding author:** Y. Suchy; Email: [yana.suchy@psych.utah.edu](mailto:yana.suchy@psych.utah.edu)

**Cite this article:** Suchy Y., DesRuisseaux L.A., Gereau Mora M., Brothers S.L., & Niermeyer M.A. (2024) Conceptualization of the term “ecological validity” in neuropsychological research on executive function assessment: a systematic review and call to action. *Journal of the International Neuropsychological Society*, 30: 499–522, <https://doi.org/10.1017/S1355617723000735>

© The Author(s), 2024. Published by Cambridge University Press on behalf of International Neuropsychological Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



**Figure 1.** The figure illustrates the increase of the usage of the term “ecological validity” in peer-reviewed articles pertaining to neuropsychological assessment.

collection method to tasks and skills required in the free and open environment” (p. 93) and “veridicality” refers to the test’s ability to “predict phenomena in the . . . ‘real world’” (p. 93)<sup>1</sup>. However, this conceptualization appears to have morphed over the years to confound EV with face validity. For example, Burgess and colleagues stated plainly in one of their publications that tests that are a “formalized version of real-world activity” are “inherently ecologically valid” (Burgess et al., 1998, p. 547), and later publications (Alderman et al., 2003; Zartman et al., 2013) echoed this sentiment, suggesting that empirical examination of highly face-valid tests’ associations with functional outcomes is not necessary.

In light of the growing number of studies that use the term EV (see Fig. 1), along with different authors using EV to refer to different concepts, it is critical for our field to gain improved insight into existing conceptualizations of EV in neuropsychological research. This line of inquiry is not new within the broader field of psychology (Araújo et al., 2007; Dunlosky et al., 2009; Schmuckler, 2001). Indeed, Holleman et al. (2020) described the term EV as being “shrouded in both conceptual and methodological confusion.” To address the need for a better understanding of EV, Pinto et al. (2023) conducted a literature review with the stated goal of examining how the term EV is defined in articles on neuropsychological assessment. While the authors confirmed that the two most-commonly used concepts in defining EV are verisimilitude and veridicality (referred to by the authors as representativeness and generalizability, respectively)<sup>2</sup>, this work had several limitations. First, the authors did not characterize the degree of agreement or disagreement among reviewed articles, or the presence of any potential misconceptions about EV, leaving the question about inconsistency among definitions unanswered. Relatedly, from among the 83 reviewed articles, only 50 were cited in the portion of results that pertained to the definition of EV,

leaving unclear how EV was defined or conceptualized among reviewed articles that were *not* cited in this section (i.e., the remaining 33 articles). Consequently, no conclusions can be drawn from Pinto et al. (2023) review about the relative frequency of different conceptualizations of EV within the literature. Second, the Pinto et al. (2023) review only included publications that used the term EV in their title, thereby excluding many relevant articles. And third, the scope of included articles was very broad, with some falling well outside of neuropsychology in general and neuropsychological assessment in particular, making it difficult to determine how the results pertained to any one conceptually homogeneous research area.

To address these limitations, we conducted a systematic review of studies that used the term EV specifically as pertaining to *neuropsychological assessment*. Given that, within neuropsychology, EV is most often discussed in the context of assessment of executive functions<sup>3</sup> (EFs; e.g. Barkley, 2012; Chaytor et al., 2006; Cripe, 1996; Manchester et al., 2004; Salimpoor & Desrocher, 2006; Wood & Bigler, 2017), we focused our review on articles that used the term EV in conjunction with tests of EF, excluding articles pertaining to other neurocognitive domains<sup>4</sup>. This allowed us to keep the cognitive construct of interest constant, thus affording greater consistency across methodologies and operationalizations pertaining to EV. Lastly, given that the term EV is heavily intertwined with the development of novel, more face-valid or naturalistic tests, we limited our review to articles that examined EV of such tests. By doing so, we were also able to examine how EV is conceptualized when pertaining to tests that are potentially characterized by both veridicality (i.e., association with real-world functioning) and verisimilitude (i.e., similarity to the real world),

<sup>1</sup>The concepts of veridicality and verisimilitude overlap with, and are sometimes referred to as, “generalizability” and “representativeness,” respectively (e.g., Bulzacka et al., 2016; Verdejo-García & Pérez-García, 2007).

<sup>2</sup>Although Pinto et al. (2023) explicitly state that generalizability and representativeness are broader than veridicality and verisimilitude, respectively, the definitions they offer for each pair of terms appear to reflect very similar concepts. Thus, the exact nature of the stated differences between the terms is not clear. Regardless, this level of nuance is beyond the scope of the present manuscript.

<sup>3</sup>Executive functions refer to those abilities that allow one to plan, organize, and successfully execute purposeful, goal-directed, and future oriented actions, thereby being critical for execution of many daily activities, such as instrumental activities of daily living.

<sup>4</sup>The disproportionate focus of EV research on EF is readily evident from any cursory review of the literature. For example, a search in PsychInfo (with “ecological validity” and individual major neurocognitive domains appearing in the article title and the words “test, measure, or instrument” appearing in the abstract) yielded 20 articles on executive functions, 11 articles on memory, 10 articles on attention, and one article on processing speed. Notably, multiple articles under memory and attention actually focused on working memory, which falls under the umbrella of EF.

**Table 1.** Inclusion and exclusion criteria

General Inclusion Criteria	Specific Inclusion Criteria	Specific Exclusion Criteria (any one suffices)
<ul style="list-style-type: none"> <li>• Peer-reviewed</li> <li>• Available in English</li> <li>• Empirical study (not a review article, call to action, etc.)</li> <li>• Not a case study or a case series</li> <li>• Participants are 6 years of age or older<sup>a</sup></li> <li>• “Ecologic*” appears in title, abstract, or key words</li> </ul>	<ul style="list-style-type: none"> <li>• Studies that considered validity, utility, feasibility, and/or psychometric properties of a test.</li> <li>• Test of interest must be “face-valid”<sup>b</sup></li> <li>• Test of interest must be designed to, or assumed to, measure EF.</li> <li>• The test must be described as either being, or having been designed to be, ecologically valid (or otherwise described as “ecological” or “ecologically informed” or other similar terms).</li> </ul>	<ul style="list-style-type: none"> <li>• Examination of self-report measures <i>only</i></li> <li>• <i>Purely</i> aiming to compare and/or validate a virtual reality task against a real-world version of the same task; put differently, validation of a <i>VR platform</i> as opposed to validation of a <i>test</i>.</li> <li>• The test is used <i>purely</i> as vehicle for answering an unrelated research question (e.g., intervention efficacy, neuroanatomic correlates, development of norms, structure of cognition, etc.) without any further discussion of the test itself or issues of ecological validity</li> <li>• Ecological validity is mentioned in the article only incidentally<sup>c</sup>.</li> </ul>

<sup>a</sup>Age limit was imposed since assessment of preschoolers, and methods and concepts surrounding such assessment, differ considerably from those associated with assessment of adults and school-age children. <sup>b</sup>The term “face-valid” is taken to mean a test that was specifically designed to resemble tasks or demands encountered in people’s life outside of the laboratory. Such tests are also at times described as “naturalistic.” <sup>c</sup>Incidental usage of the term EV refers to usage of the term outside of the goals/purposes of a given study, as adjudicated by agreement among authors.

that is, the two characteristics formally proposed as defining EV (Franzen & Wilhelm, 1996; Pinto et al., 2023).

Across reviewed articles, we aimed to examine the following questions: (1) whether the term EV was defined, and, if so, what the components of such a definition were (i.e., verisimilitude only, veridicality only, both verisimilitude and veridicality, or other notions); (2) if the term was not defined, whether there was an implied conceptualization that could be gleaned from the study design, interpretation, or justification for referring to a test as being ecologically valid; and (3) whether the usage of the term EV varied by publication year, journal’s aims and scope, test type, and study purpose.

## Method

This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021). No human data were used in preparation of this article, making the article exempt from review by University of Utah Institutional Board, and in compliance with the Helsinki Declaration.

### Data sources, search strategy, and inclusion/exclusion criteria

Comprehensive search was conducted on February 18, 2023, in MEDLINE and PsychINFO Databases<sup>5</sup>. No date limits were set on either end. Search terms were as follows: (TI ecologic\* OR AB ecologic\* or KW ecologic\*)<sup>6</sup> AND neuropsychol\* AND (executive functio\* OR executive dysfunction OR executive abilit\*). These terms were intended to target articles that used the term “ecological” (or some variant thereof) and that examined validity of face-valid and/or naturalistic tests of EF in the context of neuropsychological assessment. See Table 1 for inclusion and exclusion criteria.

<sup>5</sup>We did not search databases that are outside of the typical scope of clinical neuropsychology (e.g., engineering, history), since we were interested in how the term EV is used in neuropsychological literature. In other words, we did not search databases that do not contain collections that are relevant to neuropsychology.

<sup>6</sup>TI = title, AB = abstract, KW = keywords; for search terms with no delimiters, the term was searched in the entire body of the articles.

### Study selection

All retrieved article abstracts were first screened for *general* inclusion criteria and, when possible, *specific* inclusion and exclusion criteria (see Table 1) by the first author (YS). The remaining articles were retrieved and read by the first author (YS) and *independently* by at least one of four coauthors (LAD, MGM, SLB, MAN) to ascertain that inclusion and exclusion criteria listed in Table 1 were met. Whenever discrepancies occurred, these were adjudicated via discussion between the first author and at least one other coauthor.

### Data extraction

Data were extracted by two authors (YS reviewed all studies, and MN, SLB, LAD, and MGM each independently reviewed a subset of studies). Discrepancies between authors were adjudicated via a discussion between the first author and at least one coauthor, such that either (a) a perfect agreement was reached, or (b) the datapoint was coded “unspecified.” We extracted two types of data: (1) Information pertaining to the *conceptualization* of the term EV, and (2) relevant *correlates* (i.e., publication year, journal type, test type, and study aims) of the conceptualization of EV.

#### Conceptualizations of the term EV

To code how EV was conceptualized, we conducted a deductive content analysis (using paragraphs as units of analysis). As a first step, we generated three categories, based on a recent literature review (Pinto et al., 2023): (1) veridicality, (2) verisimilitude, and (3) other notions (see Table 2 for descriptions). Next, each article was read by at least two authors and coded for presence of text reflecting any of the three categories. In the absence of text from any of the categories, or when agreement could not be reached among coders, the article was coded as “unspecified.” See Table 3 for coding rules. For transparency, representative examples of coded statements within each article were extracted and are provided in relevant tables offering an overview of all included articles.

#### Correlates of the conceptualizations of the term EV

To examine temporal trends of the usage of the term EV, we extracted the publication year. Years were clustered into five five-year blocks. To examine differences of usage based on journal type, test types, and study aims, we used coding rules outlined in Table 4.

**Table 2.** Definition and operationalization of veridicality, verisimilitude, and “other notions” as evidence of ecological validity

Criterion for EV	Examples of Acceptable Definitions	Examples of Operationalizations of “Real-World” (and its synonyms)
Veridicality	<ul style="list-style-type: none"> <li>The ability to predict, or generalize to, functioning in the real world</li> <li>The ability to predict, or generalize to, executive difficulties in daily life</li> <li>The ability to generalize from performance in the laboratory to functioning in the real world</li> </ul>	<ul style="list-style-type: none"> <li>Self- or informant-reports of IADL functioning</li> <li>Self- or informant-report of executive lapses in daily life</li> <li>Observation of performance of IADL (or other daily) tasks</li> <li>Actual functional outcome such as functional dependency on others, occupational status, driving record, record of medication mismanagement, etc.</li> </ul>
Verisimilitude	<ul style="list-style-type: none"> <li>The resemblance to the types of tasks performed in the real-world</li> <li>A test that is designed to mimic the unstructured environment and the demands of daily functioning, including meta-tasking, prospective memory, planning, and organization</li> </ul>	<ul style="list-style-type: none"> <li>The test is an actual IADL task, such as cooking or shopping.</li> <li>The test is a virtual version of an IADL task</li> <li>The test is a pencil-and-paper version of an actual activity encountered in daily life, such as planning a party, placing medications in a pillbox, or performing mock administrative (or other similar work-related) tasks</li> </ul>
Other notions	<ul style="list-style-type: none"> <li>Sensitivity to deficits</li> <li>Construct or concurrent validity</li> </ul>	<ul style="list-style-type: none"> <li>Ability to detect differences between groups (patients vs. control)</li> <li>Evidence of impaired performance in a clinical population</li> <li>Correlation with other tests of EF</li> <li>Correlation with other tests of cognition</li> </ul>
Unspecified	<ul style="list-style-type: none"> <li>The coders were not able to identify any statements pertaining to why a test was described as being ecologically valid, or the statements were sufficiently unclear that the coders could not confidently arrive at an agreed upon interpretation.</li> </ul>	

## Results

### Search results and general description of reviewed articles

Once duplicates were eliminated, the initial search yielded 514 articles. Screening of abstracts eliminated 406 articles based on criteria in Table 1. For the remaining 108 records, full articles were reviewed, resulting in the removal of an additional 18 articles (see Fig. 2), with 90 articles included in the review<sup>7</sup>. Figure 3 illustrates that selected articles spanned 25 years from 1998 to 2022, came primarily from neuropsychology or neurorehabilitation journals, focused primarily on paper-and-pencil tests or tests that utilized computer or VR technology, and typically specified ecological or other validation as study purpose.

### Formal definitions

Only 28 articles (31%) provided a formal EV definition<sup>8</sup>. Table 5 lists these 28 studies, and Table 6 lists the 28 definitions. As seen in Table 6, 14 of the 28 definitions (50%) subscribed to the two-pronged conceptualization of EV (i.e., requiring that a test have *both* veridicality *and* verisimilitude), 10 (36%) required only that a test predict real-world functioning (i.e., veridicality), and four (14%) required only that a test appear like the real world (i.e., verisimilitude). The one characteristic that was most prevalent among the definition (24 of 28, or 86%) was that the test be able to predict a real-world functional outcome (i.e., veridicality). Of the four articles that required only verisimilitude, one (Torralva et al., 2012) also noted the ability to discriminate between groups as an additional characteristic of EV, and one (Suchy et al., 2022) implied

<sup>7</sup>Initial review of the 108 articles yielded 96 concordant inclusion/exclusion classifications, and 12 classifications in which raters felt uncertain and requested a group discussion (for the purpose of Kappa calculation, these 12 articles were considered discordant; Kappa = .617). The 12 articles were discussed among three authors and unanimous agreement was achieved.

<sup>8</sup>Initial review of articles yielded 77 concordant and 13 discordant decisions (Kappa = .63) about whether a definition was present in a given article, but unanimous agreement was achieved following a discussion between at least two authors. Additionally, initial review of the 28 definitions yielded 25 concordant and three discordant interpretations (Kappa = .799), but full agreement was achieved following a discussion among at least two authors.

that veridicality was another characteristic of EV but did not explicitly state so in the definition itself. Interestingly, provided definitions did not seem to serve as a conceptual framework for the study in all cases. Specifically, Chevignard et al. (2009), whose explicit definition included the requirement to predict functional outcome, continued to refer to the test in question as ecologically valid despite having failed to find associations between the test and two different measures of daily functioning.

### Informal usage

We next examined how EV was conceptualized among the remaining 62 articles that did not provide a definition. Notably, six of these articles fairly consistently steered away from saying that tests in question were ecologically *valid*, describing them instead as “ecologically sensitive,” “ecologically informed,” or “ecologically relevant,” or as “ecological” tests, tasks, tools, measures, or assessments. While such phrasing is more circumspect by avoiding claims of “validity,” it is also open to interpretation; for example, the term could be used *not* to imply a psychometric property, but rather as a descriptor of an overt characteristic of the test or the environment in which it was performed. Thus, these articles were excluded from further examination of the informal usage of EV. See Table 7 for an overview of these articles. For the remaining articles, we examined the explicit or implied meaning of the term EV.

### Articles that linked the results to conclusions about a test’s EV<sup>9</sup>

Table 8 provides an overview of 30 articles that did not provide a definition, but that explicitly linked conclusions about the test’s EV to the study results, thereby offering direct evidence about how EV was operationalized. From these, 18 (60%) cited prediction of functional outcome (i.e., veridicality) *alone* as evidence of the test’s EV, and an additional five (17%) cited prediction of functional outcome *in combination* with the test’s appearance (i.e., verisimilitude; Allain et al., 2014; Burgess et al., 1998; Gilboa

<sup>9</sup>Initial review of articles yielded 25 concordant and five discordant decisions (Kappa = .717) about the authors’ conceptualization of EV, but full agreement was achieved following a discussion among at least two authors.

**Table 3.** Rules for coding conceptualization of ecological validity

Code	Coding Rules
Veridicality (see Table 2)	<ul style="list-style-type: none"> <li>• Veridicality was listed as a component of a definition of EV               <ul style="list-style-type: none"> <li>○ Sample text: EV refers to the test's ability to predict daily functioning</li> </ul> </li> <li>• Veridicality was listed as a method for examining the test's EV               <ul style="list-style-type: none"> <li>○ Sample text: EV was tested by examining the test's correlation with a measure of daily functioning</li> </ul> </li> <li>• Veridicality was listed as evidence of a test's EV, either based on the study's own results or based on prior research               <ul style="list-style-type: none"> <li>○ Sample text: Previously reported correlation between the test and measure of daily functioning provide support for the test's EV</li> </ul> </li> <li>• Veridicality was linked to EV while discussing other tests or the EV concept generally               <ul style="list-style-type: none"> <li>○ Sample text: Because this test has EV, it was expected to outperform traditional tests in predicting daily functioning</li> </ul> </li> </ul>
Verisimilitude (see Table 2)	<ul style="list-style-type: none"> <li>• Verisimilitude was listed as a component of a definition of EV               <ul style="list-style-type: none"> <li>○ Sample text: EV refers to the test's similarity to the real world</li> </ul> </li> <li>• Verisimilitude was listed as a method for examining the test's EV               <ul style="list-style-type: none"> <li>○ Sample text: EV was tested by examining participants' impressions of the test's similarity to the real world</li> </ul> </li> <li>• Verisimilitude was listed as evidence of a test's EV, either based on the study's own results or based on prior research               <ul style="list-style-type: none"> <li>○ Sample text: In prior research, participants rated the test as highly similar to real-world tasks, providing support for the test's EV</li> </ul> </li> <li>• Verisimilitude was listed as evidence of a test's EV, based on the test's overt characteristics               <ul style="list-style-type: none"> <li>○ Sample text: The test closely mimicked everyday life, thereby exhibiting high EV</li> </ul> </li> <li>• Verisimilitude was linked to EV while discussing other tests or the EV concept generally               <ul style="list-style-type: none"> <li>○ Sample text: Traditional tests have poor EV due to their lack of similarity to daily life.</li> </ul> </li> </ul>
Other notions (see Table 2)	<ul style="list-style-type: none"> <li>• Other notions were listed as a component of a definition of EV               <ul style="list-style-type: none"> <li>○ Sample text: EV refers to the test's sensitivity to impairment</li> </ul> </li> <li>• Other notions were listed as a method for examining the test's EV               <ul style="list-style-type: none"> <li>○ Sample text: EV was tested by examining the test's correlation with other measures of executive functions</li> </ul> </li> <li>• Other notions were listed as evidence of a test's EV, either based on the study's own results or based on prior research               <ul style="list-style-type: none"> <li>○ Sample text: Previously reported correlation between the test and measure of executive functions provide support for the test's EV</li> </ul> </li> <li>• Other notions were linked to EV while discussing other tests or the EV concept generally               <ul style="list-style-type: none"> <li>○ Sample text: These tests do a better job discriminating between patients and controls because they are ecologically valid.</li> </ul> </li> </ul>
Unspecified	<ul style="list-style-type: none"> <li>• While EV and the above-listed notions are discussed in close proximity to each other and a linkage between them may be implied, the intention to communicate that EV is defined by such notions cannot be ascertained from the sentence structure               <ul style="list-style-type: none"> <li>○ Sample text: This ecologically valid test evidenced high similarity to daily life.</li> </ul> </li> <li>• No explanation of the term EV is provided throughout the article</li> </ul>

et al., 2019) or in combination with the test's ability to discriminate between patients and controls (Kallweit et al., 2020; Montgomery, Hatton, Fisk, Ogden, & Jansari, 2010). The remaining seven articles (23%) did *not* rely on the prediction of functional outcomes as evidence of EV. Instead, three (10%) relied on appearance (i.e., verisimilitude), either alone (Orkin Simon et al., 2022) or in combination with tests' associations with other measures (Doherty et al., 2015; Jovanovski et al., 2012); and four articles (13%) based their conclusions about EV on tests' associations with other tests of EF (La Paglia et al., 2012, 2014; Raspelli et al., 2011), or a correlation between the virtual and the real versions of the same test (Laloyaux et al., 2014). Interestingly, three articles examined correlations between the test and functional outcome but did *not* link the results of these procedures to EV; instead, these articles used alternative "validity" terminology, referring to convergent (Gilboa et al., 2019; Kenworthy et al., 2020) and concurrent validity (Orkin Simon et al., 2022).

#### Articles that used the term EV without a definition or linkages to results<sup>10</sup>

Lastly, we examined 26 remaining articles that described the tests of interest as ecologically valid but did not provide a definition or use their results as evidence of EV (see Table 9). Among these, seven articles (27%) seemed to judge EV based on test appearance (i.e., verisimilitude) alone, one (4%) relied on tests' associations with functional outcomes (i.e., veridicality) alone, and six (23%)

appeared to rely on both veridicality and verisimilitude. For the remaining 12 articles (46%), the presumed characteristics of EV could not be determined.

#### Summary of conceptualization of EV

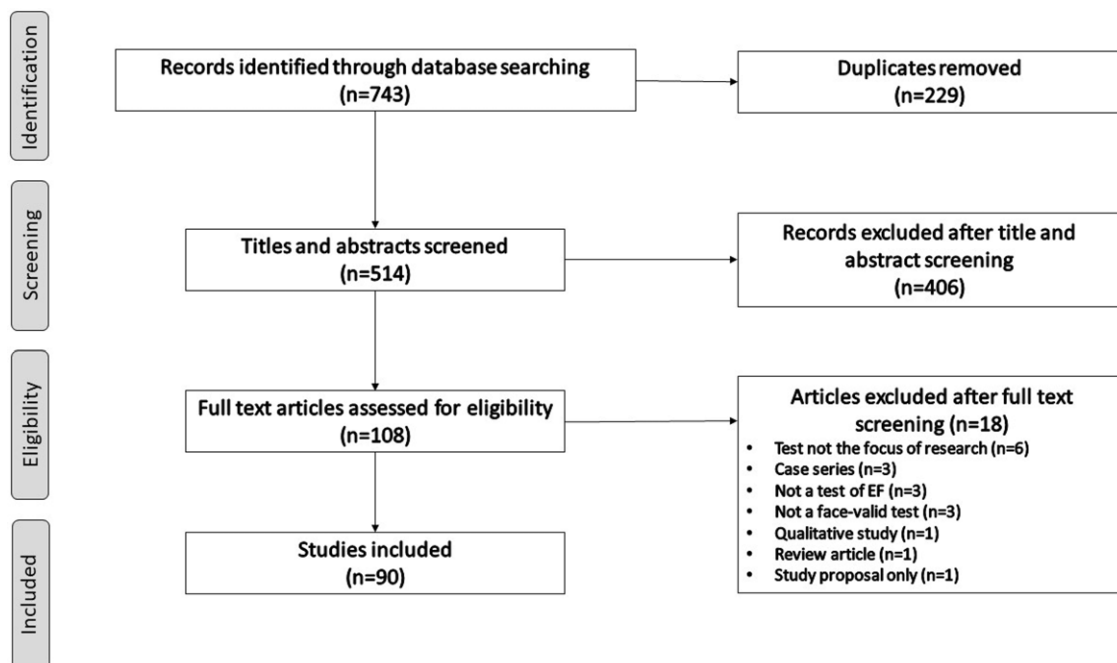
The pie chart in Figure 4 provides a summary of conceptualizations across all studies, illustrating that about two-thirds of the articles subscribed to one of the two "classic" conceptualizations of EV (i.e., either veridicality *alone*, or veridicality *together with* verisimilitude). This also means that for about one-third of the studies, the definition was comprised either of verisimilitude alone, or some combination of other notions (e.g., associations with other tests or ability to discriminate between diagnostic groups), or the meaning was unclear. The bar graph in Figure 4 illustrates that the conceptualization of EV differed dramatically [Likelihood Ratio (8) = 71.63,  $p < .001$ , Cramer's  $V = .62$ ] based on whether (a) an article provided a definition of EV and (b) if not providing a definition, whether it attempted to draw linkages between the results of the study and the test's EV. Specifically, the overwhelming majority of articles that provided a definition conceptualized EV as a test's ability to predict functional outcomes, either by this notion alone or in conjunction with tests' appearance. This was also the case (although to a lesser extent) for articles that, without providing a definition, drew some linkages between their results and the test's EV. However, about a quarter of these articles also seemed to confound EV with other notions, such as sensitivity to group differences or associations with other tests. For studies that described the tests

<sup>10</sup>Initial review of articles yielded 19 concordant and seven discordant decisions (Kappa = .682) about the authors' conceptualization of EV. After a discussion among at least two authors, agreement among raters was reached on all remaining articles.

**Table 4.** Rules for coding of correlates

Variable	Category	Inclusion Criteria
Journal Area <sup>1</sup>	Neuropsychological Neurorehabilitation Medicine & Psychiatry	Some form of “neuropsychology” appeared in journal name Some form of “rehabilitation” or rehabilitation subdiscipline appeared the journal name Some form of “medicine,” “psychiatry,” or medicine/psychiatry subdiscipline appeared in journal name
	Populations & Disorders	Name of a specific disorder, a category of disorders, or a population appeared in the journal name
	Technology	Some form of the word “technology” or technological methodology/product appeared in the journal name
	Other	This category comprised journals with broad scope that could encompass several of the above categories, or journals that could not be classified using the above criteria
	Test Type	Mock or real environment
Test Type	Computer or virtual reality technology	Tests that were performed within immersive or non-immersive virtual environment
	Paper and pencil	Tests that could be performed in a typical office/laboratory while seated at the testing table.
	Study Goals <sup>2</sup>	Examine ecological validity
Study Goals <sup>2</sup>	Examine other types of psychometric properties and/or the test’s clinical utility	Explicit statement that one of study goals was to examine any type of validity (other than ecological) or reliability, or to examine feasibility/utility of a test with a particular population or within a particular setting.
	Compare test to other tests	Explicit statement that one of the goals was to examine how a test compared to other tests with respect to any type of a number of outcomes
	Examine a population	Explicit statement that one of the study goals was to examine cognition/functioning of a particular population, with the test being used to answer such population-specific questions.
	Unclear	Study goals could not be gleaned

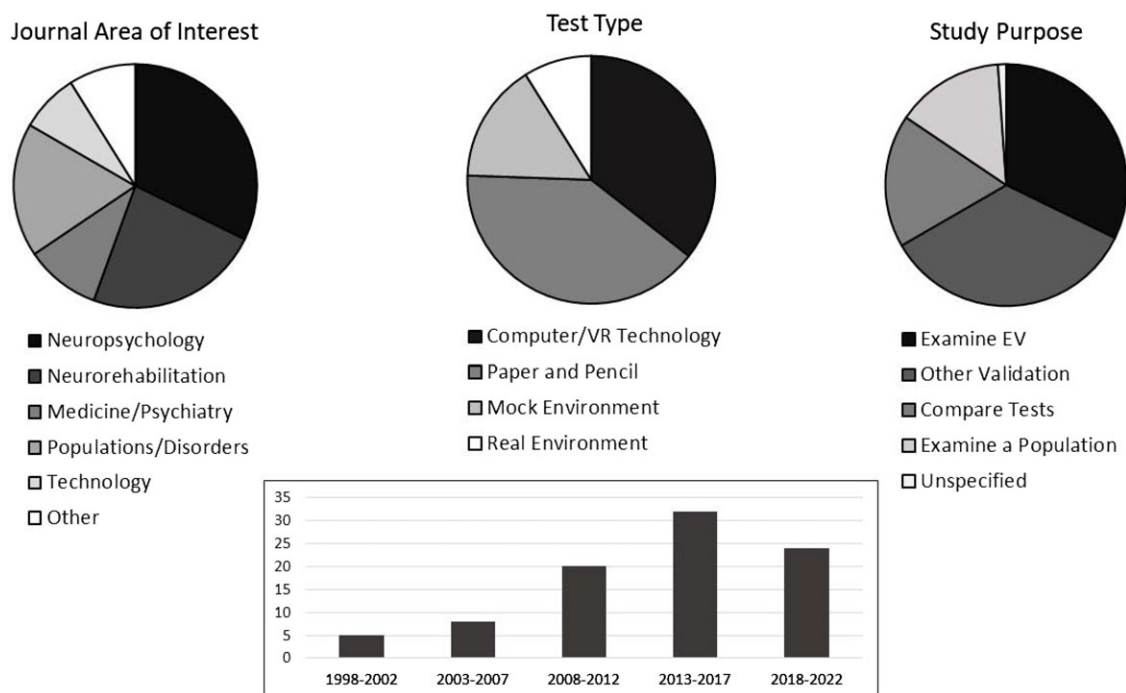
<sup>1</sup>Journal aims and scope were examined to ensure that the journal names corresponded to the primary areas of interest. If a discrepancy between journal name and aims and scope occurred, aims and scope was given precedence. For journal names that straddled two different categories, journal was classified based on aims and scope.<sup>2</sup>For the purpose of statistical analysis, studies were classified as though having only a single goal. Thus, if a study listed goals from multiple categories, the categories were ranked in the above-listed order and the highest-ranked category was used. This allowed us to categorize studies based on the degree to which ecological validity, or validity in general, was the focus of the study.

**Figure 2.** Article selection flowchart.

in question as ecologically valid without providing a definition and without drawing linkages between the study results and tests’ EV, the EV conceptualization was unclear in nearly half the cases. The remainder was about evenly split between relying purely on test appearance, or test appearance in conjunction with prediction of functional outcome.

### Correlates of usage of the term EV

As seen in Figure 5, both the usage of a definition and the conceptualization of EV varied based on study purpose [Likelihood Ratio (3) = 11.85,  $p = .008$ , Cramer’s  $V = .36$ , and Likelihood Ratio (12) = 34.92,  $p < .001$ , Cramer’s  $V = .37$ , respectively]. Specifically, studies that focused on comparison to other



**Figure 3.** The figure provides an overview of the general characteristics of 90 articles included in the present systematic review.

tests were more likely to use a definition. Additionally, studies that explicitly aimed to examine a measure's EV overwhelmingly viewed prediction of functional outcome (i.e., veridicality) as evidence of EV, whereas studies that focused on examining a population were most likely to rely on test appearance (i.e., verisimilitude) alone. Conceptualization further varied by publication year [Likelihood Ratio (16) = 28.25,  $p = .030$ , Cramer's  $V = .29$ ] and test type [Likelihood Ratio (8) = 18.65,  $p = .017$ , Cramer's  $V = .30$ ]. Specifically, as seen in Figure 6, in the first five years of the study of EV, prediction of functional outcome was invariably viewed as an aspect of EV, typically in combination with test appearance, which became less common in later years. Additionally, VR and computer tests were the most likely to rely on nontraditional definitions of EV.

## Discussion

Criticisms of inconsistent and confusing usage of the term EV in psychological research have been repeatedly raised (Araújo et al., 2007; Dunlosky et al., 2009; Holleman et al., 2020; Schmuckler, 2001). The present review examined how the term EV is used specifically in the context of neuropsychological research of novel, face-valid tests of EF. The key findings are that (a) EV is infrequently defined and (b) both formal definitions and informal usage of EV vary considerably. These findings suggest that the literature on EV of face-valid EF tests is unclear and potentially highly misleading, consistent with similar concerns raised within the broader field of psychology (Araújo et al., 2007; Dunlosky et al., 2009; Holleman et al., 2020; Schmuckler, 2001). Indeed, the present review reveals that a statement in a study's abstract or conclusions section claiming that the results supported a test's EV could be referring to different notions, including that the test: (a) predicted daily functioning, (b) differentiated clinical groups, (c) correlated with other cognitive measures, and/or (d) has face validity. This inconsistency in conceptualization, together with the frequent

absence of a formal definition, is further compounded by the fact that readers themselves likely interpret statements about EV through the lens of their own understanding of what the term means, potentially drawing highly skewed conclusions about implications for clinical practice.

## Trends over time, test types, journals, and study purpose

As illustrated in Figure 1, the usage of EV within neuropsychological publications has grown more than 20-fold over the past 25 years. With this increase in usage, there has been a drift in how the term is conceptualized. On the one hand, our results suggest that, initially, the term appeared to be exclusively taken to mean that a given test predicted functioning in daily life (i.e., veridicality), either alone or in conjunction with test appearance (i.e., verisimilitude). On the other hand, results suggest that in the past 20 years, in the literature on face-valid tests of EF, researchers have begun to rely on test appearance alone to claim EV. Additionally, wholly erroneous conceptualizations have also begun to emerge, conflating EV with sensitivity to brain injury (Torralva et al., 2012), tests' ability to differentiate groups (Kallweit et al., 2020; Montgomery et al., 2010), or construct or concurrent validity evidenced by associations with other tests (Doherty et al., 2015; Jovanovski et al., 2012; La Paglia et al., 2012, 2014; Laloyaux et al., 2014; Raspelli et al., 2011). Importantly, for some authors, EV appears to have become completely decoupled from prediction of functional outcome, as some studies that examined the association between the test and functional outcome failed to draw any connection between their results and EV (Alderman et al., 2003; Chevignard et al., 2010; Chicchi Giglioli et al., 2021; Finnanger et al., 2022; Júlio et al., 2019; Laloyaux et al., 2014; Longaud-Valès et al., 2016; Moriyama et al., 2002; O'Shea et al., 2010; Oliveira et al., 2016; Orkin Simon et al., 2022; Verdejo-García & Pérez-García, 2007; Zartman et al., 2013). Notably, some authors even claimed evidence of EV in face of their own negative findings about

**Table 5.** Overview of studies that provided an explicit definition of the term EV

Authors	Participants	Face-valid test	Journal	Study Purpose
Alderman et al. (2003)	<i>n</i> = 50, ABI <i>n</i> = 46, controls	Multiple Errands Test (MET)- Simplified Version (1)	Journal of the International Neuropsychological Society (1)	Examine EV
Bulzacka et al. (2016)	<i>N</i> = 100, Schizophrenia	MET (1)	Psychiatry Research (3)	Compare to other tests
Chevignard et al. (2009)	<i>n</i> = 10, TBI <i>n</i> = 18, controls	Children's Cooking Test (1)	Developmental Neurorehabilitation (2)	Compare to other tests
Chicchì Giglioli et al. (2021)	<i>n</i> = 18, AUD <i>n</i> = 23, controls	Virtual Cooking Test (2)	Cyberpsychology, Behavior, and Social Networking (5)	Compare to other tests
Chicchì Giglioli et al. (2018)	<i>N</i> = 354, healthy adults	EXPANSE (2)	PLoS ONE (6)	Other psychometrics
Clark et al. (2000)	<i>n</i> = 35, ADHD only <i>n</i> = 38, ADHD and ODD/ CD <i>n</i> = 11, ODD/CD only <i>n</i> = 26, controls	BADS Six Elements Test (3)	Journal of Abnormal Child Psychology (4)	Examine a population
De Oliveira et al. (2015)	<i>N</i> = 36, older adults (6 = pilot only)	ECO-VR (2)	The Spanish Journal of Psychology (6)	Other psychometrics
Engel-Yeger et al. (2009)	<i>N</i> = 208 healthy children	BADS-C (3)	Neuropsychological Rehabilitation (2)	Other psychometrics
Finnanger et al. (2022)	<i>n</i> = 75, pABI <i>n</i> = 59, controls	Children's Cooking Task (1)	Frontiers in Human Neuroscience (6)	Other psychometrics
Kourtesis et al. (2021)	<i>n</i> = 18, gamers <i>n</i> = 23, non-gamers	Virtual Reality Everyday Assessment Lab (2)	Journal of the International Neuropsychological Society (1)	Examine EV
Lamberts et al. (2010)	<i>n</i> = 35, brain injury <i>n</i> = 57, controls	Executive Secretarial Task (1)	Journal of Clinical and Experimental Neuropsychology (1)	Other psychometrics
Longaud-Valès et al. (2016)	<i>n</i> = 21, childhood frontal tumor <i>n</i> = 42, controls	BADS-C (3)	Neuropsychological Rehabilitation (2)	Other psychometrics
Moriyama et al. (2002)	<i>n</i> = 22, alcoholic <i>n</i> = 15, controls	BADS (3)	Alcoholism: Clinical and Experimental Research (4)	Compare to other tests
Nadler Tzadok et al. (2022)	part 1: <i>n</i> = 42, TBI <i>n</i> = 47, controls part 2: <i>n</i> = 28, TBI	Internet-Based Bill-Paying Task (2)	American Journal of Occupational Therapy (2)	Examine EV
Nir-Hadad et al. (2017)	<i>n</i> = 19, stroke <i>n</i> = 20, controls	Adapted Four-Item Shopping Task (2)	Neuropsychological Rehabilitation (2)	Examine EV
Norris & Tate (2000)	<i>n</i> = 36, neurological disorder <i>n</i> = 37, controls	BADS (3)	Neuropsychological Rehabilitation (2)	Examine EV
O'Shea et al. (2010)	<i>n</i> = 29, bipolar <i>n</i> = 29, controls	BADS (3)	Journal of Affective Disorders (4)	Examine a population
Rand et al. (2009)	<i>n</i> = 9, stroke <i>n</i> = 20, healthy young adults <i>n</i> = 20, healthy older adults	Virtual-MET (1)	Neuropsychological Rehabilitation (2)	Examine EV
Romundstad et al. (2022)	<i>n</i> = 74, pABI <i>n</i> = 60, controls	BADS-C (3)	Neuropsychological Rehabilitation (2)	Examine EV
Rosetti et al. (2018)	<i>N</i> = 106, children with ADHD	Ball Search Field Task (1)	Journal of Clinical and Experimental Neuropsychology (1)	Compare to other tests
Spitoni et al. (2018)	<i>n</i> = 62, anorexia nervosa <i>n</i> = 70, controls	BADS (3)	Psychiatry Research (3)	Compare to other tests
Suchy et al. (2022)	<i>N</i> = 50, older adults	Night Out Task (1)	Journal of Clinical and Experimental Neuropsychology (1)	Compare to other tests
Torralva et al. (2012)	<i>n</i> = 19, BD <i>n</i> = 15, controls	MET-Hospital Version (1)	Bipolar Disorders (4)	Compare to other tests
Tyson et al. (2008)	<i>n</i> = 36, schizophrenia <i>n</i> = 15, controls	BADS (3)	International Journal of Psychiatry in Clinical Practice (3)	Other psychometrics
Verdejo-García & Pérez-García (2007)	<i>n</i> = 37, substance dependent <i>n</i> = 37, controls	BADS (3)	Drug and Alcohol Dependence (4)	Compare to other tests
Werner et al. (2009)	<i>n</i> = 30, MCI <i>n</i> = 30, controls	Virtual Action Planning Supermarket (2)	Neuropsychological Rehabilitation (2)	Other psychometrics
Wood & Liossi (2006)	<i>N</i> = 60, severe brain injury	BADS (3)	Archives of Clinical Neuropsychology (1)	Examine EV
Ziemnik & Suchy (2019)	<i>N</i> = 50, older adults	Pillbox Test (3)	Psychological Assessment (6)	Compare to other tests

ABI = acquired brain injury, pABI = pediatric ABI, ADHD = Attention-Deficit/Hyperactivity Disorder, ASD = Autism Spectrum Disorder, AUD = Alcohol Used Disorder, BADS = Behavioral Assessment of Dysexecutive Syndrome, BADS-C = BADS for Children, MET = Multiple Errands Test, OCD = Obsessive Compulsive Disorder, ODD/CD = Oppositional Defiant Disorder/Conduct Disorder, JEF-C = Jansari assessment of Executive Functions for Children.

Codes for Journal and Test categories are presented in parentheses; journals codes: neuropsychological = 1, neurorehabilitation = 2, medicine & psychiatry = 3, populations/disorders = 4, technology = 5, and other = 6; Test type codes: mock or real environment = 1, computer or virtual environment = 2, and paper and pencil = 3.



**Table 6.** Definitions of ecological validity used in reviewed articles

Authors	Definition	Veridicality (Functional Outcome)	Verisimilitude (Appearance)
Alderman et al., (2003, p. 3132) <sup>a</sup>	... test performance is a good predictor of problems with planning and “intentionality” (i.e., goal-directed behavior) in everyday life (e.g., Burgess et al., 1998), in other words the test has good “ecological validity.” [Test] is inherently ecologically valid [if it is] a formalized version of an activity in which people naturally indulge (in Western society).	x	x
Bulzacka et al., (2016, p. 393)	The term “ecological validity” refers to the extent to which a tool is able to represent real-life cognitive problems and predict cognitive outcomes.	x	
Chevignard et al., (2009, p. 77)	In this regard, ecological validity has been defined by Sbordone (1996) as the ‘functional and predictive relation between the patient’s behavior on a set of neuropsychological tests and the patient’s behavior in a variety of real world settings’. It is concerned with: (i) the similarity of the test to behaviors required in the natural environment and (ii) the extent to which neuropsychological test results validly predict specific functioning in the natural environment of the individual (Franzen & Wilhelm, 1996; Silver 2000).	x	x
Chicchi Giglioli et al., (2021, p. 2)	“Ecological validity refers to the ability of a test to predict the individual’s real-life performance”	x	
Chicchi giglioli et al. (2018, p. 2)	“Ecological validity refers to the ability of a test to predict the individual’s real-life performance”	x	
Clark et al., (2000, p. 412)	“Ecological validity of a neuropsychological test encompasses both a theoretical relation between the test and everyday behavior, and a conceptual similarity between task demands and the demands of everyday life (Franzen & Wilhelm, 1996)”	x	x
ODe Oliveira et al., (2016, p. 157)	The term “ecological validity” refers to the functional and predictive relationship between performance on neuropsychological assessments and impairments in daily living tasks (Chaytor & Schmitter-Edgecombe, 2003; Jovanovski et al., 2012	x	
Engel-Yeger et al., (2009, p. 663)	The ecological validity of measuring EF refers to the conditions under which generalizations can be made from controlled experiments to natural real-life scenarios (Norris & Tate, 2000; Tupper & Cicerone, 1990).	x	
Finnanger et al., (2022, p. 3)	Ecological validity refers to the degree to which performance on standardized tests corresponds to the actual behavior of interest in natural settings (Barkley, 1991; Chaytor et al., 2006; Franzen and Wilhelm, 1996), and should therefore have characteristics similar to a naturally occurring behavior and to be able to predict everyday function (Franzen and Wilhelm, 1996).	x	x
Kourtesis et al., (2021, p.181)	Verisimilitude and veridicality are the two predominant approaches for achieving the ecological validity of neuropsychological tests (Franzen & Wilhelm, 1996; Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006).	x	x
Lamberts et al., (2010, p. 56)	Whilst there have been a number of definitions of ecological validity, in a neuropsychological context it has been defined by Sbordone (1996) as: “the functional and predictive relationship between the patient’s performance on a set of neuropsychological tests and the patient’s behavior in a variety of real-world settings” (p. 16)	x	
Longaud-Vales et al. (2016, p. 561)	ecological validity ... refers to the conditions under which generalization can be made from controlled conditions to real-life scenarios (Gioia & Isquith, 2004; Silver, 2000).	x	
Moriyama et al., (2002, p.1239)	Two general aspects of ecological validity have been stressed. The first is verisimilitude, or the similarity of the data collection method to tasks and skills required in a free and open environment. The second is veridicality, or the extent to which test results reflect or can predict phenomena in the open environment or “real world” (Franzen and Wilhelm, 1996).	x	x
Nadler Tzadok et al., (2022, p.4)	When the findings of an assessment can explain daily functioning or behavior in the participant’s own environment, it is considered to have ecological validity (Portney & Gross, 2020).	x	
Nir-Hadad et al., (2017, p. 810)	Parsons (2011, 2015) suggested that the extent to which a given virtual environment will achieve ecological validity will depend on its verisimilitude and veridicality such that the tasks performed within it correspond to key aspects of real-world activities and environments, and provide outcome measures relevant to the practical problem being investigated.	x	x
Norris & Tate (2000, p. 34)	Ecological validity refers to the conditions under which generalizations can be made from controlled experimental situations to naturalistic environments (Tupper & Cicerone, 1990).	x	
O’Shea et al., (2010, p. 337)	Ecologically valid cognitive tests have been developed to simulate everyday situations that appear to be troublesome for patients with even mild cognitive impairment. These tests map cognitive demands in the everyday environment and have greater face validity than standard tests (Norris and Tate, 2000; Chaytor and Schmitter-Edgecombe, 2003).		x
Rand et al., (2009, p. 585)	According to Burgess et al. (2006) such assessments have ecological validity and should show a degree of “representativeness” (the association between the results of a clinical test to situations encountered beyond laboratory conditions) and “generalisability” (the extent to which poor performance on a given test will predict problems outside the laboratory) (p.195).	x	x

(Continued)

Table 6. (Continued)

Authors	Definition	Veridicality (Functional Outcome)	Verisimilitude (Appearance)
Romundstad (2022, p. 2)	... there is a need to develop EF tests with an effort to provide increased representativeness and generalizability to real-life situations (i.e., ecological validity; Burgess et al., 2006).	x	x
Rosetti et al., (2018, p. 576)	In general terms ecological validity refers to the capacity to extrapolate from clinical or experimental contexts to those of real-world situations (Bronfenbrenner, 1977, 1979; Schmuckler, 2001).	x	
Spitoni et al., (2018, p. 284)	Ecological validity is the ability to generalize the results of controlled experiments to natural occurring events in everyday life. With regard to EFs, ecological validity can be described as the “functional and predictive relationship between the patient’s behavior on a set of neuropsychological tests and the patient’s behavior in a variety of real-world settings...” (Sbordone, 1996). Thus, an ecologically valid test has features that are similar to naturally occurring behaviors and has value in predicting everyday function (Franzen and Wilhelm, 1996).	x	x
Suchy et al., (2022, p. 1) <sup>b</sup>	Ecological validity, as initially conceptualized, refers to the need for tests to tap into the same processes (or have the same demands) as those that are needed for daily functioning (Sbordone & Long, 1996).		x
Torralva et al., (2012, p. 119) <sup>c</sup>	... ecological validity – that is, they actually mimic real-life scenarios and have proven to be sensitive in the detection of frontal dysfunction in various neurological and psychiatric conditions (Torralva et al., 2009, 2012; Roca et al., 2008, 2010).		x
Tyson et al., (2008, p. 48)	Chaytor and Schmitter-Edgecombe (2003) have proposed that ecological validity is established through verisimilitude and veridicality.	x	x
Verdejo-Garcia & Perez-Garcia (2007, p. 48)	Ecologically valid assessments should address two main aspects: (i) representativeness, and (ii) generalizability (Franzen and Wilhelm, 1996; Chaytor and Schmitter-Edgecombe, 2003). The former refers to the extent to which a clinical test corresponds in form and context to a situation encountered outside the laboratory; the latter refers to the degree to which performance on the test will be predictive of problems outside the laboratory (Burgess et al., 2006).	x	x
Werner et al., (2009, p. 302)	The term ecological validity which refers to the representativeness of a task is central in the area of EF assessment because these are closely associated with real-life complex situations (such as shopping or preparing a meal) that require organization and structuring (Burgess et al., 2000).		x
Wood & Lioffi (2006, p. 429)	Chaytor and Schmitter-Edgecombe (2003) propose two concepts upon which ecologically valid tests rely, verisimilitude and veridicality. Verisimilitude reflects the degree to which a test resembles the cognitive demands of real life tasks. The related concept of veridicality refers to the degree to which performance on a neuropsychological test is empirically related to measures of everyday functioning.	x	x
Ziemnik & Suchy (2019, p. 2)	Ecological validity in the context of neuropsychological evaluations refers to the relationship between performance on neuropsychological tests administered in controlled office settings and actual behavioral outcomes in the real world (Franzen & Wilhelm, 1996). There are two qualities that are considered when establishing ecological validity. These are (a) veridicality, which is present when performance on the test correlates with measures of real-world functioning, or, more preferably, with the real-world functioning itself and (b) verisimilitude, which is present when the demands of the task resemble the demands encountered in everyday environments (Franzen & Wilhelm, 1998).	x	x

A “definition” was defined as an *explicit and direct* explanation of the meaning or components of “ecological validity.” Indirect allusions or meaning that were only implied were not considered definitions.

<sup>b</sup>Phrasing of the definition appears to suggest that either veridicality or verisimilitude is sufficient.

<sup>b</sup>Article implies that verisimilitude is also a characteristic of EV, but this is not included in the definition as worded here.

<sup>c</sup>Definition suggests that sensitivity to brain damage is also needed in addition to verisimilitude.

veridicality (Chevignard et al., 2009; Clark et al., 2017; Gilboa et al., 2019). Taken together, these results illustrate that the usage of the term EV has become increasingly inconsistent, departing further from original conceptualizations (Franzen & Wilhelm, 1996; Sbordone, 1996). That said, as seen in Figure 6, the past decade evidences an apparent trend toward returning to the original two-pronged conceptualization, perhaps as a function of emerging criticisms of confusing usage (Araújo et al., 2007; Dunlosky et al., 2009; Holleman et al., 2020; Schmuckler, 2001).

Interestingly, usage also varied by study purpose. First, studies that focused primarily on comparisons of the utility of various tests were more likely to provide a formal definition of EV (Figure 5), likely because the comparisons were typically made between tests that were presumed to be ecologically valid and those that were

not<sup>11</sup>. Thus, provision of a definition was necessary to justify grouping of tests into ecological vs. non-ecological categories. Additionally, studies that set out to empirically examine tests’ EV were most likely to associate EV with predictions of functional outcomes, likely because examination of EV necessitated explicit operationalization of the term and explicit hypotheses. In contrast, studies that focused on particular disorders or populations tended to rely on test appearance (i.e., verisimilitude) as evidence of EV. This may be explained by the fact that articles that focus on a given population may not necessarily be interested in *prediction* of outcomes, but rather may be more focused on *characterizing*

<sup>11</sup>Although the characteristics that are associated with the notion of EV (i.e., veridicality and verisimilitude) typically reflect continuous variables, EV is frequently treated as a dichotomy in the reviewed literature.

**Table 7.** Overview of studies that did not provide a definition and did not use the full term “ecological validity” when describing tests

Authors	Participants	Face-valid test	Journal	Study Purpose
Canali et al. (2007)	<i>n</i> = 17, early AD <i>n</i> = 17, controls	BADS (3)	Dementia & Neuropsychologia (4)	Examine a population
Chevalère et al., (2013)	<i>N</i> = 20, Prader-Willi Syndrome	BADS (3)	Journal of Applied Research in Intellectual Disabilities (4)	Examine a population
Espinosa et al. (2009)	<i>n</i> = 50, MCI <i>n</i> = 50, mild AD <i>n</i> = 50, controls	BADS (3)	Journal of the International Neuropsychological Society (1)	Other psychometrics
Laloyaux et al. (2013)	<i>n</i> = 21, bipolar disorder <i>n</i> = 21, controls	Computerized Shopping Task (2)	Psychiatry Research (3)	Other psychometrics
Pishdadian et al. (2022)	<i>n</i> = 25, schizophrenia <i>n</i> = 32, controls	Computerized Breakfast Task (2)	Applied Neuropsychology (1)	Compare to other tests
Toussaint-Thorin et al. (2013)	<i>N</i> = 13, children with developmental dyspraxia	Children’s Cooking Task (1)	Annals of Physical and Rehabilitation Medicine (2)	Compare to other tests

ABI = acquired brain injury, AD = Alzheimer’s Disease, BADS (3) = Behavioral Assessment of Dysexecutive Syndrome, BD = bipolar disorder, MCI = mild cognitive impairment, pABI = pediatric acquired brain injury, TBI = traumatic brain injury.

Codes for Journal and Test categories are presented in parentheses; journals codes: neuropsychological = 1, neurorehabilitation = 2, medicine & psychiatry = 3, populations/disorders = 4, technology = 5, and other = 6; test type codes: mock or real environment = 1, computer or virtual environment = 2, and paper and pencil = 3.

patients’ functioning. In this type of research, naturalistic tests of EF may then be assumed to provide an insight into patients’ daily lives, thereby representing an outcome rather than a predictor. Thus, it is understandable that high verisimilitude represents the most salient and valued aspect of EV in this line of research.

Additionally, considerable differences in conceptualization of EV were also evident by test type. Specifically, research on paper-and-pencil tests and tests administered in real or mock environments linked EV primarily to prediction of functional outcomes (either alone or in conjunction with test appearance), whereas research on tests performed in virtual or computer environments tended to equate EV with test appearance or with other nontraditional notions. It is likely that the latter is related to the fact that developers of computer-based naturalistic environments focus primarily on ensuring that such tests sufficiently approximate the natural environment, and in the process perhaps lose sight of the principal reason for test development, that is, the test’s eventual clinical utility.

### Pitfalls associated with the term ecological validity

#### Clinical misconceptions

As the present review shows, there is no clear consensus about the meaning of the term EV, resulting in considerably inconsistent use across studies. This, in and of itself, is not all that unusual. Other terms used in neuropsychological literature are similarly plagued by the lack of a universally-accepted definition, with the neuro-cognitive domain of EF representing a salient example (Suchy, 2015). However, there is a critical difference between the problems with conceptualization of EF and conceptualization of EV. Specifically, inconsistencies in EF conceptualization pertain to certain discrete disagreements, such as whether the term is unitary or multidimensional, or how broad the umbrella of EF should be. Aside from these differences of opinion, there are core EF abilities that are fairly universally agreed upon, and differences in definitions are not likely to have a meaningful impact on how study results are interpreted or applied in clinical practice. In contrast, despite some overlap in definitions and usage of the term EV, differences in definitions appear to lead to diametrically opposed and mutually inconsistent interpretations and conclusions, with potentially clinically meaningful ramifications.

A clear example of diametrically opposed conclusions can be gleaned from studies that apply the term EV to traditional

measures of EF. Specifically, consistent with the veridicality interpretation of EV, a number of studies that have found association between traditional EF tests (i.e., tests with *low* verisimilitude) and functional outcomes have explicitly concluded that, based on their findings, such test are *ecologically valid* (e.g., Chiu et al., 2018; García-Molina et al., 2012; Hoskin et al., 2005; Kibby et al., 1998; Lea et al., 2021; Mitchell & Miller, 2008; Odhuba et al., 2005; Possin et al., 2014; Reynolds et al., 2019; Silverberg et al., 2007; Sudo et al., 2015; Van der Elst et al., 2008; Ware et al., 2012). Yet, it is fairly common for articles that focus on novel face-valid tests to claim, as a matter of unequivocal fact, that traditional EF tests *lack ecological validity* (e.g., Allain et al., 2014; Chevignard et al., 2008; Jovanovski et al., 2012; La Paglia et al., 2012; Longaud-Valès et al., 2016; Renison et al., 2012; Rosetti et al., 2018; Shimoni et al., 2012; Torralva et al., 2012; Valls-Serrano et al., 2018; Verdejo-García & Pérez-García, 2007; Werner et al., 2009). While these latter statements are sometimes meant to simply communicate the tests’ lack of *face validity* (or, potentially, a failure to tap into all cognitive domains needed for daily functioning), they often also communicate (explicitly or implicitly) that these tests are *not able* to predict functional outcomes. Indeed, even if the authors do not purposely intend to comment on the test’s ability to predict outcomes, such conclusions may be drawn by readers, based on their own idiosyncratic ways of conceptualizing EV. Figure 7 illustrates how the slippage between the veridicality and verisimilitude notions of EV leads to a deductive fallacy with erroneous conclusions that contradict research findings and potentially impact clinical practice. Indeed, clinicians may favor tests with greater face validity over traditional measures, regardless of the strength of empirical evidence (or lack thereof) about such novel tests’ ability to predict functional outcomes.

#### Psychometric misconceptions

Interestingly, in the present review, even among the articles that did examine associations between a measure and a functional outcome as evidence of EV, some nevertheless strongly implied that the test characteristics were sufficient to describe the test as ecologically valid. For example, Alderman et al. (2003) stated that a test is “inherently ecologically valid” if it resembles real-world tasks (p. 37); and Zartman et al. (2013) followed suit, stating that the typical criticism of traditional tests’ ability to predict IADLs “does not apply” (p. 316) to their novel face-valid test, implying that such tests *can be assumed* to predict real-world functioning. From this

**Table 8.** Overview of studies that did not provide a definition but did link their results to conclusions about a test's ecological validity

Authors	Participants	Face-valid test	Journal	Study Purpose	EV Conceptualization and Sample Text
Allain et al. (2014)	<i>n</i> = 24, AD <i>n</i> = 32, controls	Non-Immersive Virtual Coffee Task (2)	Journal of the International Neuropsychological Society (1)	Examine EV	<b>VR:</b> We found significant relations between ... virtual scores and IADL scale score, thereby supporting the ecological validity of the [test]. (p474) <b>VS:</b> tasks may enhance ... ecological validity by presenting the patient with functional situations that resemble daily life. (p. 470)
Bertens et al. (2016)	<i>n</i> = 60, healthy adults	Modified Six Elements Test (3)	Applied Neuropsychology (1)	Examine EV	<b>VR:</b> the ecological validity of the [test] was evaluated using ... a self-rating questionnaire measuring everyday executive performance. (p. 35)
Besnard et al. (2016)	<i>n</i> = 19, ABI <i>n</i> = 19, controls	Non-Immersive Virtual Coffee Task (2)	Applied Neuropsychology (1)	Examine EV	<b>VR:</b> Ecological validity of test results was assessed by comparing them to significant others' reports on patients' occupational activities, interpersonal relationships, and independent living skills. (p. 224)
Burgess et al. (1998)	<i>n</i> = 92, neurological <i>n</i> = 216, controls	Simplified Six Elements Test (3)	Journal of the International Neuropsychological Society (1)	Examine EV	<b>VR:</b> Broadly implied throughout, but no clear explicit statement made. <b>VS:</b> there are executive tests that are inherently "ecologically valid" since they are little more than formalized versions of real-world activities (p. 547)
Chevignard et al. (2008)	<i>n</i> = 45, ABI <i>n</i> = 12, controls	Children's Cooking Test (1)	Neuropsychological Rehabilitation (2)	Examine a population	<b>VR:</b> ... the [test of everyday EF problems] was also correlated with the total number of errors in the [test], indicating good ecological validity (p.476)
Clark et al. (2017)	<i>n</i> = 15, ABI <i>n</i> = 16, controls	Baycrest MET (1)	Journal of Abnormal Child Psychology (4)	Examine EV	<b>VR:</b> ... several studies have indicated that ecological validity of executive function assessments can be established via informant- or clinician-report (p 679)
Cuberos-Urbano et al. (2013)	<i>N</i> = 30, ABI	MET-Hospital Version (1)	Journal of Clinical and Experimental Neuropsychology (1)	Examine EV	<b>VR:</b> ... results further support the ecological validity of the [test] by demonstrating that patients' performance on the test can predict the behavioral problems ... in everyday life ... (p.334)
Doherty et al. (2015)	<i>n</i> = 46, healthy adults (lifespan)	Computer-Based Cooking Task (2)	Frontiers of Behavioral Neuroscience (6)	Other psychometrics	<b>VS:</b> ... The core components of the real-life task should be captured by the ecologically valid version ... (p ??) <b>COR:</b> ... the relationships found between [test] indices and standardized measures hold ... promise for ... the [test] as an ecologically valid measure .. (p. ??).
Fisher et al. (2022)	<i>n</i> = 52, adolescents with ADHD <i>n</i> = 50, controls	Weekly Calendar Planning Activity - Hebrew Middle/High School Version (3)	The American Journal of Occupational Therapy (2)	Examine EV	<b>VR:</b> ... ecological validity [was] supported by significant associations with measures of [participation in daily activities] (p. 6)
Gilboa et al. (2019)	<i>n</i> = 29, ABI <i>n</i> = 30, controls	JEF-C (2)	Neuropsychological Rehabilitation (2)	Other psychometrics	<b>VR &amp; VS:</b> Feedback from clinicians, parents, and patients indicated the potential of [test] in terms of ecological and predictive validity: however, a study exploring the [test] ability to predict real-world functioning was not conducted. (p. 1377)
Josman et al., (2014)	<i>n</i> = 24, stroke <i>n</i> = 24, controls	Virtual Action Planning-Supermarket (2)	Journal of Stroke and Cerebrovascular Diseases (4)	Examine EV	<b>VR:</b> ... support for concurrent and ecological validity of the [test] is provided by the significant ... correlations between ... [test] ... and the [daily living] total score. (p. 884)
Jovanovski et al., (2012)	<i>N</i> = 30, healthy young adults	Multitasking in the City Test (2)	Applied Neuropsychology (1)	Examine EV	<b>VS:</b> ... the ecologically valid assessment of executive functions should involve the evaluation of common everyday behaviors. (p. 171) <b>COR:</b> The current study characterized ... [test]'s ... relationship ... with standardized cognitive tests... [This] preliminary evidence suggests that the [test] may provide an ecologically valid method ... (p. 180).
Jovanovski et al. (2012)	<i>N</i> = 30, healthy young adults	Multitasking in the City Test (2)	Applied Neuropsychology: Adult (1)	Other psychometrics	<b>VR:</b> ... the ecological validity of the [test] was demonstrated by the statistically significant correlations between the [test] and [daily living] (p. 216)

Table 8. (Continued)

Authors	Participants	Face-valid test	Journal	Study Purpose	EV Conceptualization and Sample Text
Kallweit et al. (2020)	<i>n</i> = 36, adults with ADHD <i>n</i> = 36, controls	Daily Life Task (3)	Journal of Clinical and Experimental Neuropsychology (1)	Examine EV	<b>VR:</b> ... correspondence between performance in the new tasks and functional aspects ... would serve as an important indicator of a more ecologically valid assessment ... (p. 570) <b>GRP:</b> ... if the new tasks would show more pronounced differences between the two groups ... this would reflect an aspect of ecological validity. (p. 570)
Kenworthy et al. (2020)	<i>n</i> = 129, ASD <i>n</i> = 93, ADHD	Executive Function Challenge Test (3)	Journal of the International Neuropsychological Society (1)	Other psychometrics	<b>VR:</b> It is related to parent-reported everyday EF problems, indicating its potential as an ecologically valid ... measure (p. 730)
La Paglia et al. (2014)	<i>n</i> = 30, OCD <i>n</i> = 30, controls	Virtual MET (2)	Annual Review of Cybertherapy and Telemedicine (5)	Other psychometrics	<b>COR:</b> Results of the analyses carried out within the present study [i.e., correlations with other tests] confirm previous data ... about ecological validity of [test] ... (p44)
La Paglia et al. (2012)	<i>n</i> = 10, OCD <i>n</i> = 10, controls	Virtual MET (2)	Annual Review of Cybertherapy and Telemedicine (5)	Examine a population	<b>COR:</b> ... the significant correlation found between the [test] and the neuropsychological battery support the ecological validity of [test] (p. 101)
Laloyaux et al. (2014)	<i>n</i> = 21, schizophrenia <i>n</i> = 20, controls <i>n</i> = 14, Schizophrenia	Computerized Meeting Preparation Task (2)	Psychiatry Research (3)	Examine EV	<b>COR:</b> ... the [test] was significantly correlated with a real version of the task, pointing to good ecological validity. (p. 168)
Maeir et al. (2011)	<i>N</i> = 30, ABI	MET-Hospital Version (1)	Occupational Therapy Journal of Research (2)	Examine EV	<b>VR:</b> The current results provide supporting evidence for the ecological validity of the [test] ... The [test] scores found at discharge were significantly correlated with [daily living] 3 months later. (p. S44)
Montgomery et al. (2010)	<i>n</i> = 23, ecstasy-polydrug users <i>n</i> = 26, controls	Jansari Agnew Akesson Murphy-task (2)	Human Psychopharmacology (3)	Other psychometrics	<b>VR:</b> ... such tasks may not have 'functional significance' (ecological validity) (p. 319) <b>GRP:</b> ... differences on specific subscales could reflect the better ecological validity of the [test] (p. 323)
Orkin Simon et al. (2022)	<i>N</i> = 40, healthy children	JEF-C Hebrew Version (2)	Neuropsychological Rehabilitation (2)	Other psychometrics	<b>VS:</b> Informal feedback from parents and discussions with participants indicated the potential of [test] in terms of ecological and predictive validity (p. 301)
Raspelli et al. (2011)	<i>n</i> = 9, stroke <i>n</i> = 10, healthy young adults <i>n</i> = 10, healthy older adults	Virtual-MET (1)	Studies in Health and Technology (5)	Examine EV	<b>COR:</b> Correlations between [test] and some traditional executive functions measures provide preliminary support for the ecological and construct validity of the [test] (p. 31)
Renison et al. (2012)	<i>n</i> = 30, TBI <i>n</i> = 30, controls	Virtual Library Task (2)	Journal of the International Neuropsychological Society (1)	Examine EV	<b>VR:</b> ... [test] significantly predicted everyday EF suggesting [it is] ecologically valid ... (p. 440)
Rosenblum et al. (2015)	<i>n</i> = 23, children with ADHD <i>n</i> = 24, controls	Do-Eat (1)	Neuropsychological Rehabilitation (2)	Examine EV	<b>VR:</b> ... the ecological validity of the [test] is demonstrated by the correlations found between the [test] scores and the parents' questionnaire ... (p. 414)
Roy et al. (2015)	<i>n</i> = 120, healthy children	BADS-C (3)	Journal of Clinical and Experimental Neuropsychology (1)	Examine EV	<b>VR:</b> Results of the current study prompt us to be cautious about the point of a better "ecological validity" of the [test]. In fact, children's performances ... do not appear to be accurately reflected in everyday-life questionnaires ... (p. 967)
Sanders & Schmitter-Edgecombe (2017)	<i>n</i> = 32, healthy young adults <i>n</i> = 64, healthy older adults	Amap Task (1)	Neuropsychological Rehabilitation (2)	Examine EV	<b>VR:</b> ... significant relationships between [test] and informant-report of everyday functioning highlight the ecological validity of the [test] ... (p. 759)
Schmitter-Edgecombe et al. (2021)	<i>n</i> = 26, MCI <i>n</i> = 122, healthy older adults <i>n</i> = 57, healthy young adults	Night Out Task (1)	Archives of Clinical Neuropsychology (1)	Examine EV	<b>VR:</b> To demonstrate ecological validity, a hierarchical linear regression analysis was conducted to determine whether [test] would account for significant variance in self-reported everyday functioning (p. 542-543)

(Continued)

Table 8. (Continued)

Authors	Participants	Face-valid test	Journal	Study Purpose	EV Conceptualization and Sample Text
Scott et al. (2011)	n = 60, HIV+ n = 25, controls	Multitasking Test (3)	Neuropsychology (1)	Examine EV	VR: ... the correspondence between the [test] and IADL outcome measures ... points to its potential incremental ecological relevance (i.e., validity) (p.517)
Siu & Zhou (2014)	n = 63, children with ADHD	BADS-C (3)	Journal of Child Neurology (3)	Examine EV	VR: The present study showed that the [daily living scale] was significantly correlated to only the [test], suggesting ... ecological validity ... (p. 613)
Steverson et al. (2017)	n = 60, controls N = 40, intellectual disability	Modified MET for Intellectual Disabilities (1)	Journal of Applied Research in Intellectual Disabilities (4)	Examine EV	VR: ... examine the strength of the relationship between ... [test] and [daily living scale] to assess the ecological validity of the [test]. (p. 263)

ABI = acquired brain injury, ADHD = Attention-Deficit/Hyperactivity Disorder, ASD = Autism Spectrum Disorder, BADS = Behavioral Assessment of Dysexecutive Syndrome, COR = correlation with other tests, GRP = examination of group differences, JEF-C = Jansari assessment of Executive Functions for Children, MET = Multiple Errands Test, OCD = Obsessive Compulsive Disorder, VR = Veridicality, VS = Verisimilitude.  
Codes for Journal and Test categories are presented in parentheses; journals codes: neuropsychological = 1, neurorehabilitation = 2, medicine & psychiatry = 3, populations/disorders = 4, technology = 5, and other = 6; Test type codes: mock or real environment = 1, computer or virtual environment = 2, and paper and pencil = 3.

perspective, EV (along with face validity) appears to carry a special status in that it is treated as though it is exempt from the requirement of empirical evidence. Such status, of course, contradicts the whole notion of test validation, wherein other types of validity (i.e., concurrent, predictive, construct, etc.) all require empirical confirmation. If treated in this manner, EV would then not reflect a test's psychometric property (as other types of validity do), but rather a somewhat nebulous vernacular for readily apparent and potentially clinically irrelevant test characteristics. It is perhaps for these reasons that some authors opted to avoid linking together the words "ecological" and "validity," describing their tests instead as "ecological assessments" or as "ecologically informed," and other similar variations. Notably, we have repeatedly shown that naturalistic elements of a test *do not* necessarily improve upon prediction of objective real-world outcomes beyond measures with low face validity (Suchy et al., 2022; Suchy et al., *in press*; Ziemnik & Suchy, 2019), demonstrating that predictive validity cannot be assumed based on test appearance alone.

### Communication breakdown and a call to action

Interestingly, as mentioned earlier, some studies that did find evidence for the test's ability to predict functional outcome did not link their study results with the term EV, referring instead to convergent or concurrent validity (Chevignard et al., 2010; Finnanger et al., 2022; Gilboa et al., 2019; Kenworthy et al., 2020; Oliveira et al., 2016; Orkin Simon et al., 2022; Pishdadian et al., 2022; Zartman et al., 2013). Yet, the methods and results of these articles could have legitimately warranted claims of EV (if relying on either of the two classical definitions of EV; Sbordone, 1996; Franzen & Wilhem, 1996), given that (a) the employed tests possessed face validity and (b) the tests showed the ability to predict functional outcomes. Conversely, some studies in the present review that failed to find any associations between the test and functional outcome nevertheless continued to describe their tests as ecologically valid (Chevignard et al., 2009; Clark et al., 2017; Gilboa et al., 2019), contradicting the most prevalent conceptualizations of EV, at least as evidenced in the present review. It is our position that these and other grossly contradictory claims and interpretations (also see "clinical misconceptions" above) reported throughout this review represent a highly problematic breakdown in communication, rendering the term EV essentially meaningless and potentially harmful.

To address this breakdown in communication, Holleman et al. (2020) called upon reviewers and editors to "safeguard journals from publishing papers where terms such as 'ecological validity' ... are used without specification." While we fully support this call, the present review suggests that provision of a definition may not be enough. For example, as noted earlier, Chevignard et al. (2009) continued to describe the test in question as being ecologically valid, despite their own empirical findings that contradicted the definition provided by the authors themselves. Similarly, Alderman et al. (2003) claimed that their test was inherently ecologically valid due to its high face validity, despite having provided a definition that explicitly stated that EV is defined by the test's ability to predict functional outcome. In other words, it appears that the term EV is implicitly linked *not* to empirical evidence of validity, but rather to subjective impressions about test appearance. Indeed, as reviewed above, the notion that highly naturalistic tests can be assumed to be ecologically valid has

been repeatedly propagated in the literature (Burgess et al., 1998; Zartman et al., 2013).

We therefore call upon our profession to consider “retiring” the term EV, replacing it instead with concrete and readily interpretable terminology that well predates the usage of the term EV, specifically, *criterion* validity. Criterion validity (and its components, concurrent and predictive validity) is linked to clear empirical methodology, can be readily interpreted as the test’s association with a concrete external criterion, and has clear clinical implications. Indeed, Larrabee (2015) in his prominent article on the types of validity in clinical neuropsychology acknowledged that the term EV is sometimes used as a synonym for criterion validity. Reliance on more concrete terminology would not only improve communication but might also help address the unwarranted but widely held misgivings about the utility of traditional tests of EF. These misgivings emerged alongside the emergence of the term EV and appear to be based solely on the traditional tests’ lack of verisimilitude. In response to these misgivings, our field has been committing precious resources (financial, creative, scientific) to the development of novel “naturalistic” tests, with limited evidence that such tests will improve our clinical practice. Indeed, from among the 50+ novel tests developed with the goal of improving EV, only one instrument (Behavioral Assessment of Dysexecutive Syndrome; BADS; Wilson et al., 1996) has thus far been translated into regular clinical practice (Rabin et al., 2007). Reclaiming traditional and more meaningful validation terminology would offer hope that both novel and traditional assessment approaches would focus on true empirical validation (i.e., criterion validity), in place of subjective and untested impressions about the importance of test appearance invoked by the term EV.

### Limitations

The present review has several limitations. Perhaps the most salient is the fact that for some articles our interpretation of the conceptualization of EV was based on somewhat subjective impressions of what the authors intended to communicate. Specifically, although a number of articles provided explicit statements about how EV was conceptualized, many were much less explicit. For such articles, the interpretation of what the authors intended to communicate could potentially vary somewhat from one reader to the next. To assuage this problem, all articles were read independently by two raters, and disagreements among raters were adjudicated via a discussion between at least two coauthors. Ultimately, there were no instances where discrepancies could not be resolved. Importantly, the first author (YS) participated in all such adjudications, assuring that the same set of principles was applied evenly to all decisions. Additionally, to ensure that we did not over-interpret vague statements about the concept of EV, 12 articles were coded as “unspecified” about how EV was conceptualized. Relatedly, our deductive approach imposed the notions of verisimilitude and veridicality, thereby potentially overlooking subtle nuances within those concepts. That said, it is unlikely that such subtle nuances could have been reliably coded for each article, and, if they could be coded, the results would undoubtedly demonstrate an even greater divergence of opinions about the conceptualization of EV. In other words, by taking a more molar approach (i.e., collapsing across subtle differences in conceptualization of verisimilitude and veridicality), we employed fairly conservative criteria for divergence of conceptualizations. Thus, it is noteworthy that even with such conservative approach gross differences in conceptualization of EV emerged. Lastly, along

the same lines, it is possible that a different group of authors would arrive at a different classification scheme for journal type, test type, and study purpose. While we acknowledge this possibility, the current scheme is a result of extensive thoughtful discussions among all authors and is clearly and transparently outlined and reported in tables.

Another potential limitation is that we only examined articles that used the term EV in the context of novel, face-valid tests of EF. This decision was made for two reasons. First, it is in this literature where the term EV is applied most frequently and where calls to action have been made for the development of new, more “ecological” tests of EF (e.g., Burgess et al., 2006; Spooner & Pachana, 2006). Second, it is in this literature that the term EV is most likely to be conflated with face validity. Specifically, studies that carry out empirical examinations of EV of traditional tests of EF overwhelmingly use the term EV to imply the ability to predict functional outcome. This is understandable, since the absence of face validity of traditional tests is self-evident; thus, if EV were conceptualized as reflecting face validity, empirical examination would be irrelevant. In other words, such studies cannot be conflating EV with face validity, potentially rendering a review unnecessary. That said, it is still possible that such studies might conflate the term EV with other test characteristics, such as the ability to differentiate among groups. The present study has not addressed potential variability in how EV is conceptualized in that literature. Nevertheless, it is highly likely that inclusion of such literature would increase the number of studies that conceptualize EV as veridicality (i.e., prediction of functional outcome) alone, altering the percentages presented in the present review.

Additionally, our literature search did not include all databases. Relatedly, our search terms may have missed some articles that would have been relevant. That said, our review was quite comprehensive and included many articles not covered by the Pinto et al. (2023) review; indeed, from among the 90 articles included in our review, only nine were also included in the Pinto et al. (2023) review. This discrepancy between the two reviews is most likely a reflection of the fact that Pinto et al. (2023) only reviewed articles that included the term EV in the article title. Indeed, in the present review, the majority of reviewed articles referred to EV in the abstract, key words, and the body of the article, without mentioning it in the title. Nevertheless, a yet larger sample of articles would increase statistical power, thereby potentially affording a better insight into factors that are associated with various EV conceptualizations.

Lastly, we do not offer or recommend a particular definition or operationalization of EV, since extensive discussions of EV as a construct can be found in multiple chapters of the Sbordone and Long (1996) book, as well as in many articles and reviews published since then (e.g., Burgess et al., 2006, Chaytor et al., 2006; Spitoni et al., 2018). Additionally, our review was not intended to provide guidance on how to conduct validation research, since methodology and scientific rigor of the reviewed articles were not examined here. However, a detailed review of typical methodological pitfalls associated with ecological validation is provided in our separate systematic review (Suchy et al., *in press*), in which we recommend (a) clearly defining all terminology and steering away from the term EV unless absolutely necessary, (b) seeking convergence across multiple outcome measures, (c) controlling for relevant confounds, and (d) examining incremental validity. The latter point is particularly important for the determination of clinical utility, given that veridicality occurs on a continuum. In other words, it is not the presence or

**Table 9.** Overview of studies that did not linked study findings to an instrument’s ecological validity

Authors	Participants	Face-valid test	Journal	Study Purpose	EV Conceptualization and Sample Text
Almeida et al., (2006)	n = 50, ADHD n = 50, controls	BADS-C (3)	Neuropsychologia (1)	Examine a population	<b>VS:</b> One of the advantages that this test has is its ecological validity ... since the [test] ... is very similar to the participants’ daily activities. (p. 297)
Canali et al., (2011)	n = 41, probable mild AD n = 41, controls	BADS (3)	Revista Brasileira de Psiquiatria (2)	Compare to other tests	Unspecified
Carral-Fernández et al. (2016)	n = 32, anorexia nervosa n = 42, controls	Zoo map (3)	The Clinical Neuropsychologist (1)	Examine a population	<b>VR:</b> The [test] is superior to other tests in ecological validity in terms of predicting [daily living scale] ... (p. 231-232) <b>VS:</b> Tests with ecological validity, such as [test], have been developed to enhance the similarity between the task demands of the test and the demands imposed in the everyday environment. (p. 236)
Chevignard et al. (2010)	n = 10, mild TBI n = 15, moderate TBI n = 21, controls	Children’s Cooking Task (1)	Brain Impairment (2)	Other psychometrics	Unspecified
Cipresso et al. (2014)	n = 15, PD-MCI n = 15, PD, NC n = 15, controls	Virtual MET (2)	Frontiers in Behavioral Neuroscience (6)	Compare to other tests	Unspecified
Denmark et al. (2019)	n = 19, ABI n = 19, controls	JEF (2)	Neuropsychological Rehabilitation (2)	Compare to other tests	<b>VS:</b> The “ecological” approaches have tended to use ... real-world activity, which is time consuming...VR offers a way of creating more realistic “real world” activities within the clinic or laboratory. (p. 768)
Gamito et al. (2015)	n = 25, AUD n = 24, controls	Virtual Kitchen Test (2)	Methods of Information in Medicine (5)	Other psychometrics	<b>VR:</b> ... lack ecological validity because they propose tasks that do not effectively replicate patients’ impaired daily life activities ... (p.122)
Hill & Bird (2006)	n = 22, ASD n = 22, controls	BADS (3)	Neuropsychologia (1)	Examine a population	Unspecified
Jansari et al. (2014)	n = 6, ABI n = 6, controls n = 17, ABI n = 30, controls	JEF (2)	Brain Impairment (2)	Other psychometrics	<b>VR:</b> ... the preliminary suggestion from these findings is that [test] is an ecologically valid assessment that identifies the real-world impairments ... (p. 80). <b>VS:</b> ... [traditional tests have poor ecological validity because they] do not fundamentally resemble the actual tasks that individuals might have difficulty with. (p. 72)
Júlio et al. (2019)	n = 15, early manifest HD n = 15, premanifest HD  n = 19, controls	EcoKitchen (2)	Frontiers in Psychology (6)	Other psychometrics	<b>VR:</b> ... the relations between [test] and a real-life kitchen performance (preferably at home and not in a Lab setting) would help to corroborate its ecological validity. (p. 16) <b>VS:</b> ... to increase the realism of the task (and thus its ecological validity), known commercial brands were used to depict the foods and beverages included in the kitchen setting). (p. 6)
Klinger et al. (2004)	n = 6, PD n = 5, controls	Virtual Supermarket (2)	Annual Review of CyberTherapy and Telemedicine (5)	Examine a population	<b>VS:</b> ... advent of virtual reality technology allows the presentation of scenarios or scripts that are ecologically valid (i.e. very close to daily situations). (p. 50)
Logue et al. (2015)	N = 179, older males	Pillbox Test (3)	The Clinical Neuropsychologist (1)	Other psychometrics	Unspecified
Oliveira et al. (2016)	Pilot: N = 6, older adults Primary: N = 37, older adults	ECO-VR (2)	The Spanish Journal of Psychology (6)	Other psychometrics	<b>VS:</b> ... there still is uncertainty about the ecological validity of the traditional tests ... as they are not often linked directly to the demands of daily life activities ... (p. 1)
Poncet et al. (2015)	N = 160, ABI	Cooking Task (1)	Neuropsychological Rehabilitation (2)	Other psychometrics	Unspecified
Radomski et al. (2018)	n = 33, service members with TBI n = 50, service member controls	Change of Quarters Duty Test (1)	Military Medicine (4)	Other psychometrics	<b>VR &amp; VS:</b> Because ecologically valid ... tests typically lack the structure and organization inherent in most traditional ... tests, they may better approximate real-life demands and predict the quality of everyday functioning. (p. e215)

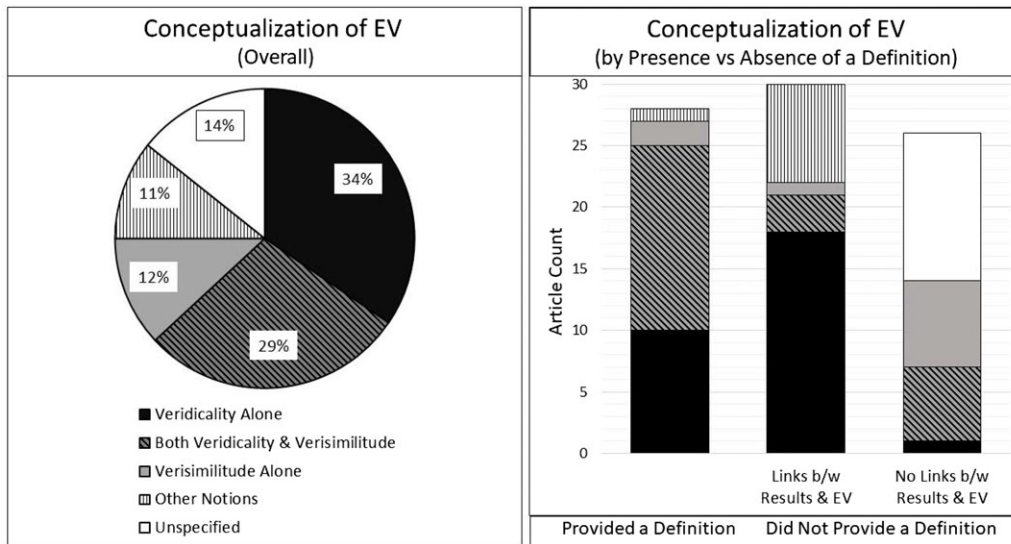


Table 9. (Continued)

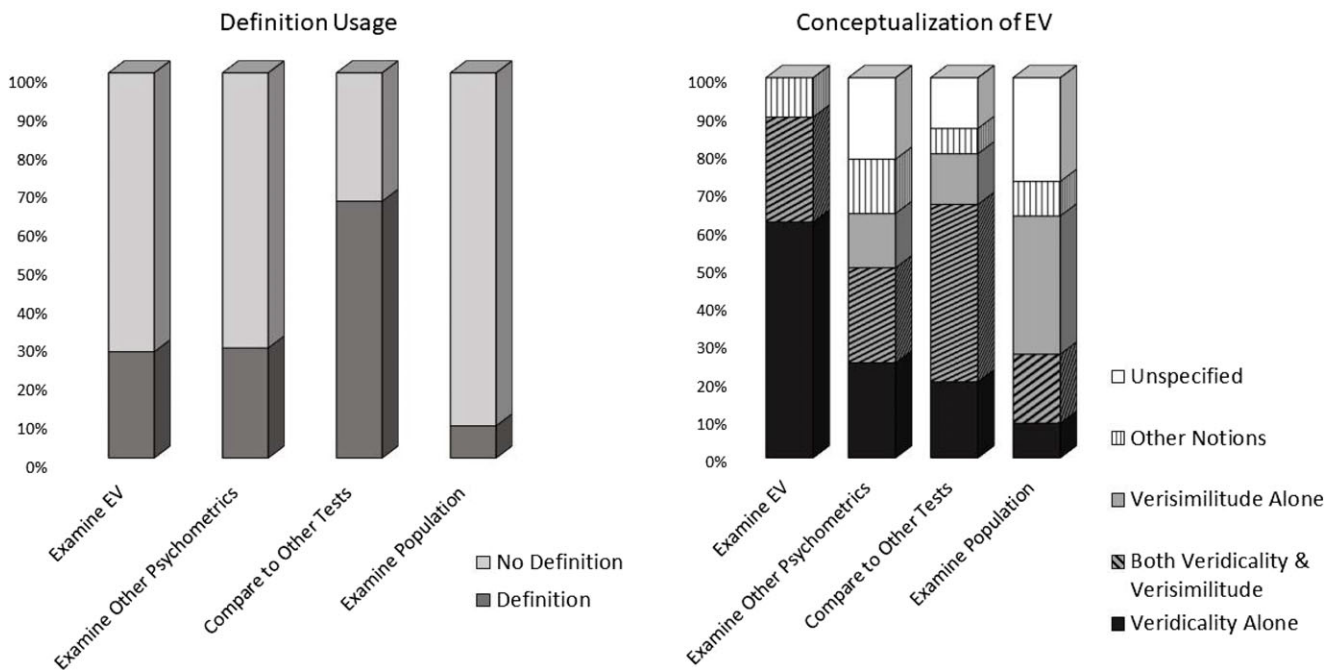
Authors	Participants	Face-valid test	Journal	Study Purpose	EV Conceptualization and Sample Text
Shimoni et al. (2012)	<i>n</i> = 25, male children with ADHD <i>n</i> = 25, controls	BADS-C (3)	Research in Developmental Disabilities (4)	Examine a population	Unspecified
Siddiqui et al. (2019)	<i>n</i> = 49, schizophrenia <i>n</i> = 55, controls	Multitasking in the City Test (2)	Schizophrenia Research (4)	Examine a population	Unspecified
Torralva et al. (2013)	<i>N</i> = 117, ADHD	The Hotel Task (3)	Journal of Attention Disorders (4)	Compare to other tests	VS: ... more “ecological” in nature, for they resemble real-life demands ... (p. 12)
Tranel et al. (2007)	<i>n</i> = 9, VMPC <i>n</i> = 8, PFC <i>n</i> = 17, nonPFC <i>n</i> = 20, controls	MET (1) Six Elements Test (3)	Journal of Clinical and Experimental Neuropsychology (1)	Examine a population	Unspecified
Valls-Serrano et al. (2018)	<i>n</i> = 60, substance use <i>n</i> = 30, controls	MET-Contextualized Version (1)	Journal of the International Neuropsychological Society (1)	Compare to other tests	VR: ... superior discriminability and ecological/predictive validity (e.g., in relation to treatment outcomes) ... (p. 348) VS: Traditional tests of executive functioning have been criticized for their low ecological validity. Shallice and Burgess (1991) developed the [test] where participants are required to perform multiple tasks in a real-world setting. (p. 347)
Webb et al. (2022)	<i>n</i> = 105, stroke <i>n</i> = 124, controls	Oxford Digital MET (2)	Neuropsychological Rehabilitation (2)	Other psychometrics	Unspecified
White et al. (2009)	<i>n</i> = 45, children with ASD <i>n</i> = 27, controls	BADS-C (3)	Autism Research (4)	Examine population	VS: ... novel “ecologically valid” tests have been designed which tap into real-life scenarios that are relevant to and representative of everyday behavior. (p. 138)
Wilson et al. (1998)	<i>n</i> = 216, controls <i>n</i> = 78, neurologic disorder <i>n</i> = 31, schizophrenia	BADS (3)	Neuropsychological Rehabilitation (2)	Other psychometrics	VR: This test was designed to predict everyday memory problems in people with brain injury. It was one of the earliest ecologically valid neuropsychological tests (p. 214-215). VS: ... using tasks analogous to those required in everyday life activities ... [referring to the battery described as ecologically valid]. (p. 215)
Wolf et al. (2017)	<i>n</i> = 14, mild stroke <i>n</i> = 20, controls	Complex Task Performance Assessment (3)	Neuropsychological Rehabilitation (2)	Other psychometrics	Unspecified
Wood & Liossi (2007)	<i>N</i> = 118, ABI	BADS-C (3)	Journal of International Neuropsychological Society (1)	Other psychometrics	Unspecified
Zartman et al. (2013)	<i>n</i> = 40, neurological <i>n</i> = 40, medical N-40, controls	Pillbox Test (3)	Archives of Clinical Neuropsychology (1)	Other psychometrics	VS: ... the [test] [has] greater ecological validity as the examiner does not function as the patient’s “frontal lobes” by employing corrective actions (p. 316)

ABI = acquired brain injury, AD = Alzheimer’s Disease, ADHD = Attention-Deficit/Hyperactivity Disorder, ASD = Autism Spectrum Disorder, AUD = Alcohol Used Disorder, BADS = Behavioral Assessment of Dysexecutive Syndrome, BADS-C = Behavioral Assessment of Dysexecutive Syndrome- Children, COR = correlation with other tests, GRP = examination of group differences, HD = Huntington’s Disease, JEF = Jansari assessment of Executive Functions, MET = Multiple Errands Test, MCI = Mild Cognitive Impairment, PD = Parkinson’s Disease, PD-NC = Parkinson’s Disease-normal cognition, TBI = Traumatic Brain Injury, VR = Veridicality, VS = Verisimilitude.

Codes for Journal and Test categories are presented in parentheses; journals codes: neuropsychological = 1, neurorehabilitation = 2, medicine & psychiatry = 3, populations/disorders = 4, technology = 5, and other = 6; test type codes: mock or real environment = 1, computer or virtual environment = 2, and paper and pencil = 3.



**Figure 4.** The figure provides an overall summary of the conceptualization of the term ecological validity (EV) across 84 articles that used the full term “ecological validity” as pertaining to a test of interest. Of note, six articles are excluded, due to reliance on less explicitly terminology (e.g., “ecological relevance” or “ecological tests”).



**Figure 5.** The figure illustrates how the explicitly stated purposes of individual studies related to whether an article provided a definition of ecological validity (EV), and to how the term EV was conceptualized. “Definition” graph is based on all 90 articles reviewed for this study. Conceptualization graph is based on 84 articles that used the full term “ecological validity.”

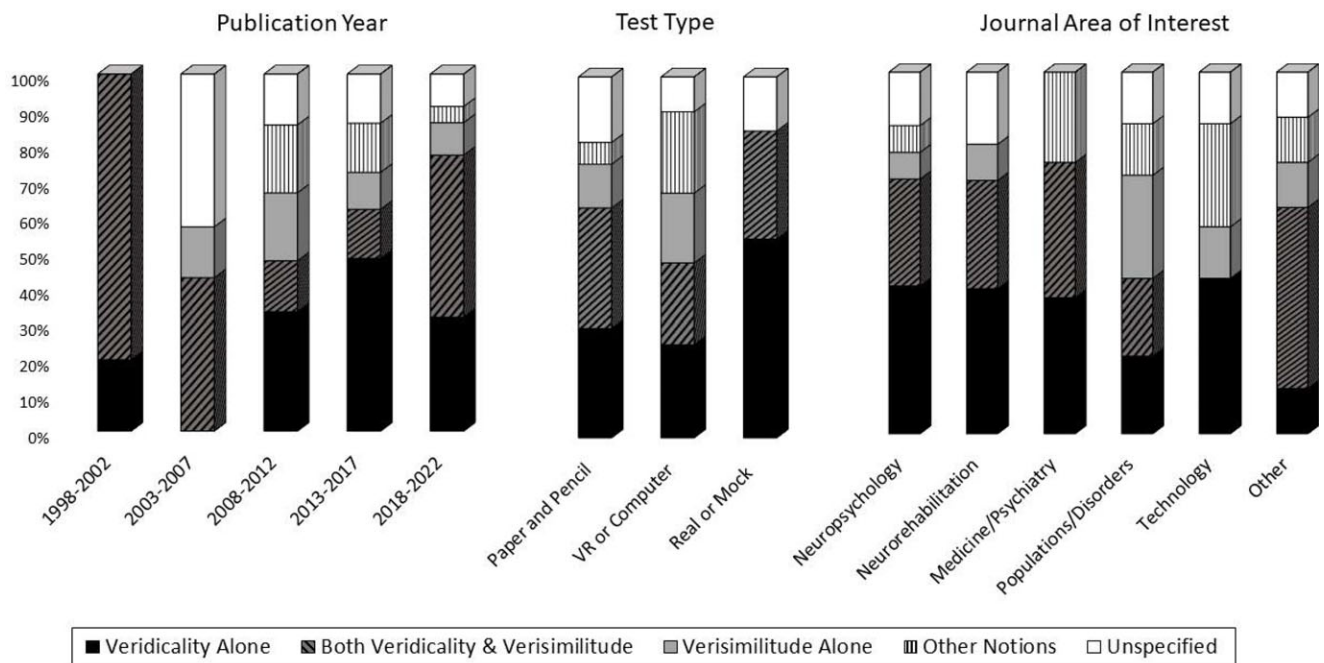
absence of veridicality that determines clinical utility, but rather the degree to which test scores predict functional outcomes above and beyond performances on other available instruments.

**Conclusions**

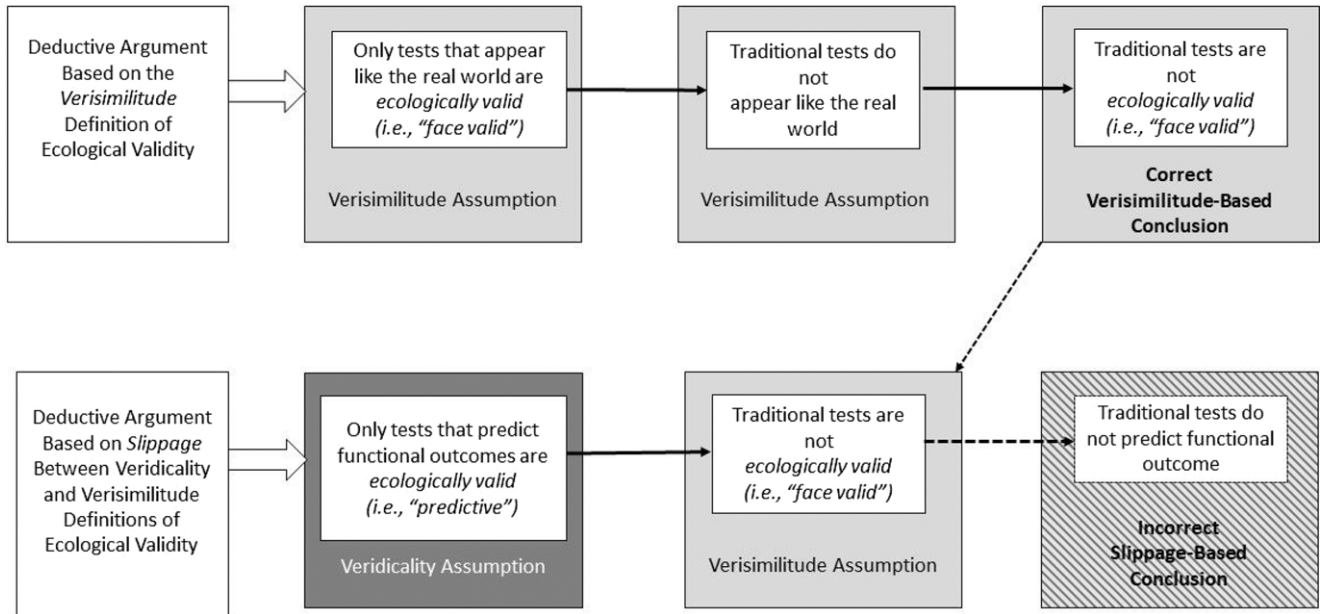
The present systematic review provides compelling evidence that the term EV is conceptualized highly inconsistently, at least in the literature on novel, face-valid or naturalistic, tests of EF. This inconsistent use is likely contributing to misconceptions about the utility of both traditional and novel instruments, potentially

harming clinical practice. Specifically, despite empirical evidence to the contrary, the permeability among different EV conceptualizations leads to the impressions that (a) novel EF tests that appear like the real world can automatically be assumed to predict daily functioning, and (b) traditional tests of EF cannot possibly predict daily functioning due to their low face validity. While we strongly support the call put forth by Holleman et al. (2020) that editors and reviewers ensure that the usage of the term EV in publications be accompanied by clear definitions and operationalizations, the present review suggests that provision of a definition may simply not be enough to remedy the pervasive breakdown in

### Conceptualization of Ecological Validity by Year, Test, and Journal



**Figure 6.** The figure illustrates the associations between how the term ecological validity was conceptualized and publication year, test type, and journal area. Differences were statistically significant for publication year and test type. Based on 84 articles that used the full term “ecological validity.” VR = virtual reality. “Real or mock”=real or mock up environments.



**Figure 7.** The figure illustrates how the slippage between the veridicality and verisimilitude conceptualizations of ecological validity can lead to logically-flawed conclusions, specifically, that traditional tests of executive functioning cannot predict functional outcomes due to their lack of verisimilitude. Extensive literature shows that this conclusion is incorrect.

communication. Therefore, we call upon our field to consider retiring the term EV and replacing it with traditional terminology, namely criterion validity, which, at least according to some authors, refers to the same concept (e.g., Larrabee, 2015).

**Financial support.** None.

**Competing interests.** None.

**References**

Alderman, N., Burgess, P. W., Knight, C., & Henman, C. (2003). Ecological validity of a simplified version of the multiple errands shopping test. *Journal of the International Neuropsychological Society*, 9(1), 31–44, Retrieved

- from. <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2003-02691-006&site=ehost-live>
- Allain, P., Foloppe, D. A., Besnard, J., Yamaguchi, T., Etcharry-Bouyx, F., Le Gall, D., Nolin, P., & Richard, P. (2014). Detecting everyday action deficits in Alzheimer's disease using a nonimmersive virtual reality kitchen. *Journal of the International Neuropsychological Society*, 20(5), 468–477. <https://doi.org/10.1017/S1355617714000344>
- Almeida, O. P., Waterreus, A., & Hankey, G. J. (2006). Preventing depression after stroke: Results from a randomized placebo-controlled trial. *Journal of Clinical Psychiatry*, 67(7), 1104–1109.
- Anisfeld, M. (1968). Disjunctive concepts? *Journal of General Psychology*, 78(2), 223–228. <https://doi.org/10.1080/00221309.1968.9710436>
- Araújo, D., Davids, K., & Passos, P. (2007). Ecological validity, representative design, and correspondence between experimental task constraints and behavioral setting: Comment on Rogers, Kadar, and Costall, 2005. *Ecological Psychology*, 19(1), 69–78. <https://doi.org/10.1080/10407410709336951>
- Barkley, R. A. (1991). The ecological validity of laboratory and analogue assessment methods of ADHD symptoms. *Journal of Abnormal Child Psychology*, 19, 149–178. <https://doi.org/10.1007/BF00909976>
- Barkley, R. A. (2012). *Executive functions: What they are, how they work, and why they evolved*. The Guilford Press. <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2012-15750-000&site=ehost-live>
- Bertens, D., Fasotti, L., Egger, J. I. M., Boelen, D. H. E., & Kessels, R. P. C. (2016). Reliability of an adapted version of the modified six elements test as a measure of executive function. *Applied Neuropsychology: Adult*, 23(1), 35–42. <https://doi.org/10.1080/23279095.2015.1012258>
- Besnard, J., Richard, P., Banville, F., Nolin, P., Aubin, G., Le Gall, D., Richard, I., & Allain, P. (2016). Virtual reality and neuropsychological assessment: The reliability of a virtual kitchen to assess daily-life activities in victims of traumatic brain injury. *Applied Neuropsychology: Adult*, 23(3), 223–235. <https://doi.org/10.1080/23279095.2015.1048514>
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, 32, 513–531. <https://doi.org/10.1037/0003-066X.32.7.513>
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design* (1st ed.). Cambridge, MA: Harvard University Press.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd edn). University California Press.
- Bulzacka, E., Delourme, G., Hutin, V., Burban, N., Méary, A., Lajnef, M., Leboyer, M., & Schürhoff, F. (2016). Clinical utility of the multiple errands test in schizophrenia: A preliminary assessment. *Psychiatry Research*, 240, 390–397. <https://doi.org/10.1016/j.psychres.2016.04.056>
- Burgess, P., Alderman, N., Forbes, C., Costello, A., Coates, L. M.-A., Dawson, D., Anderson, N., Gilbert, S., Dumontheil, I., & Channon, S. (2006). The case for the development and use of “ecologically valid” measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society*, 12(2), 194–209. <https://doi.org/10.1017/S1355617706060310>
- Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society*, 4(6), 547–558.
- Burgess, P. W., Veitch, E., de Lacy Costello, A., & Shallice, T. (2000). The cognitive and neuroanatomical correlates of multitasking. *Neuropsychologia*, 38, 848–863. [https://doi.org/10.1016/S0028-3932\(99\)00134-7](https://doi.org/10.1016/S0028-3932(99)00134-7)
- Canali, Fabiola, Dozzi Brucki, S. M., & Amodeo Bueno, O. F. (2007). Behavioural assessment of the dysexecutive syndrome (BADS) in healthy elders and Alzheimer's disease patients: Preliminary study. *Dementia & Neuropsychologia*, 1(2), 154–160. <https://doi.org/10.1590/S1980-57642008dn10200007>
- Canali, Fabiola, Brucki, S. M. D., Bertolucci, P. H. F., & Bueno, O. F. A. (2011). Reliability study of the behavioral assessment of the dysexecutive syndrome adapted for a Brazilian sample of older-adult controls and probable early Alzheimer's disease patients. *Revista Brasileira de Psiquiatria*, 33(4), 1–8. <https://doi.org/10.1590/S1516-44462011005000015>
- Carral-Fernández, L., González-Blanch, C., Goddard, E., González-Gómez, J., Benito-González, P., Bustamante-Cruz, E., & Gómez del Barrio, A. (2016). Planning abilities in patients with Anorexia Nervosa compared with healthy controls. *The Clinical Neuropsychologist*, 30(2), 228–242. <https://doi.org/10.1080/13854046.2016.1147603>
- Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, 13(4), 181–197. <https://doi.org/10.1023/B:NERV.0000009483.91468.fb>
- Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*, 21(3), 217–227. <https://doi.org/10.1016/j.acn.2005.12.002>
- Chevalere, J., Postal, V., Jauregui, J., Copet, P., Laurier, V., & Thuilleaux, D. (2013). Assessment of executive functions in Prader-Willi Syndrome and relationship with intellectual level. *Journal of Applied Research in Intellectual Disabilities*, 26, 309–318. <https://doi.org/10.1111/jar.12044>
- Chevignard, M. P., Catroppa, C., Galvin, J., & Anderson, V. (2010). Development and evaluation of an ecological task to assess executive functioning post childhood TBI: The children's cooking task. *Brain Impairment*, 11(2), 125–143. <https://doi.org/10.1375/brim.11.2.125>
- Chevignard, M. P., Servant, V., Mariller, A., Abada, G., Pradat-Diehl, P., & Laurent-Vannier, A. (2009). Assessment of executive functioning in children after TBI with a naturalistic open-ended task: A pilot study. *Developmental Neurorehabilitation*, 12(2), 76–91. <https://doi.org/10.1080/17518420902777019>
- Chevignard, M. P., Taillefer, C., Picq, C., Poncet, F., Noulhiane, M., & Pradat-Diehl, P. (2008). Ecological assessment of the dysexecutive syndrome using execution of a cooking task. *Neuropsychological Rehabilitation*, 18(4), 461–485. <https://doi.org/10.1080/09602010701643472>
- Chicchì Giglioli, I. A., de Juan Ripoll, C., Parra, E., & Alcañiz Raya, M. (2018). EXPANSE: A novel narrative serious game for the behavioral assessment of cognitive abilities. *PLoS ONE*, 13(11), e0206925. <https://doi.org/10.1371/journal.pone.0206925>
- Chicchì Giglioli, I. A., Pérez Gálvez, B., Gil Granados, A., & Alcañiz Raya, M. (2021). The virtual cooking task: A preliminary comparison between neuropsychological and ecological virtual reality tests to assess executive functions alterations in patients affected by alcohol use disorder. *Cyberpsychology, Behavior, and Social Networking*, 24(10), 673–682. <https://doi.org/10.1089/cyber.2020.0560>
- Chiu, E.-C., Wu, W.-C., Hung, J.-W., & Tseng, Y.-H. (2018). Validity of the wisconsin card sorting test in patients with stroke. *Disability and Rehabilitation: An International, Multidisciplinary Journal*, 40(16), 1967–1971. <https://doi.org/10.1080/09638288.2017.1323020>
- Cipresso, P., Albani, G., Serino, S., Pedrolì, E., Pallavicini, F., Mauro, A., & Riva, G. (2014). Virtual multiple errands test (VMET): A virtual reality-based tool to detect early executive functions deficit in parkinson's disease. *Frontiers in Behavioral Neuroscience*, 8, 405. <https://doi.org/10.3389/fnbeh.2014.00405>
- Clark, A. J., Anderson, N. D., Nalder, E., Arshad, S., & Dawson, D. R. (2017). Reliability and construct validity of a revised baycrest multiple errands test. *Neuropsychological Rehabilitation*, 27(5), 667–684. <https://doi.org/10.1080/09602011.2015.1117981>
- Clark, C., Prior, M., & Kinsella, G. J. (2000). Do executive function deficits differentiate between adolescents with ADHD and oppositional defiant/conduct disorder? A neuropsychological study using the six elements test and hayling sentence completion test. *Journal of Abnormal Child Psychology*, 28(5), 403–414.
- Cripe, L. I. (1996). The ecological validity of executive function testing. In *Ecological validity of neuropsychological testing* (pp. 171–202). Gr Press/St Lucie Press, Inc.
- Cuberos-Urbano, G., Caracuel, A., Vilar-López, R., Valls-Serrano, C., Bateman, A., & Verdejo-García, A. (2013). Ecological validity of the multiple errands test using predictive models of dysexecutive problems in everyday life. *Journal of Clinical and Experimental Neuropsychology*, 35(3), 329–336. <https://doi.org/10.1080/13803395.2013.776011>
- de Oliveira Cardoso, C., Zimmermann, N., Paraná, C. B., Gindri, G., de Pereira, A. P. A., & Fonseca, R. P. (2015). Brazilian adaptation of the hotel task: A tool for the ecological assessment of executive functions. *Dementia & Neuropsychologia*, 9(2), 156–164. <https://doi.org/10.1590/1980-57642015DN92000010>

- Denmark, T., Fish, J., Jansari, A., Tailor, J., Ashkan, K., & Morris, R. (2019). Using virtual reality to investigate multitasking ability in individuals with frontal lobe lesions. *Neuropsychological Rehabilitation*, 29(5), 767–788. <https://doi.org/10.1080/09602011.2017.1330695>
- Doherty, T. A., Barker, L. A., Denniss, R., Jalil, A., & Beer, M. D. (2015). The cooking task: Making a meal of executive functions. *Frontiers in Behavioral Neuroscience*, 9, 22. <https://doi.org/10.3389/fnbeh.2015.00022>
- Dudycha, A. L., Dumoff, M. G., & Dudycha, L. W. (1973). Choice behavior in dynamic environments. *Organizational Behavior & Human Performance*, 9(2), 328–338. [https://doi.org/10.1016/0030-5073\(73\)90056-1](https://doi.org/10.1016/0030-5073(73)90056-1)
- Dunlosky, J., Bottiroli, S., & Hartwig, M. (2009). Sins committed in the name of ecological validity: A call for representative design in education science. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 430–440). Routledge/Taylor & Francis Group, <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2010-06038-022&site=ehost-live>
- Engel-Yeger, B., Josman, N., & Rosenblum, S. (2009). Behavioural assessment of the dysexecutive syndrome for children (BADS-C): An examination of construct validity. *Neuropsychological Rehabilitation*, 19(5), 662–676. <https://doi.org/10.1080/09602010802622730>
- Espinosa, A., Alegret, M., Boada, M., Vinyes, G., Valero, S., Martínez-Lage, P., Peña-Casanova, J., Becker, J., Wilson, B., & Tárraga, L. (2009). Ecological assessment of executive functions in mild cognitive impairment and mild Alzheimer's disease. *Journal of the International Neuropsychological Society*, 15(5), 751–757. <https://doi.org/10.1017/S135561770999035X>
- Farmer, J. E., & Eakman, A. M. (1995). The relationship between neuropsychological functioning and instrumental activities of daily living following acquired brain injury. *Applied Neuropsychology*, 2(3–4), 107–115. [https://doi.org/10.1207/s15324826an0203&4\\_2](https://doi.org/10.1207/s15324826an0203&4_2)
- Finnanger, T. G., Andersson, S., Chevnard, M., Johansen, Gøril O., Brandt, A. E., Hypher, R. E., Risnes, K., Rø, T. B., & Stubberud, J. (2022). Assessment of executive function in everyday life—Psychometric properties of the norwegian adaptation of the children's cooking task. *Frontiers in Human Neuroscience*, 15, 761755. <https://doi.org/10.3389/fnhum.2021.761755>
- Fisher, O., Berger, I., Grossman, E. S., Tai-Saban, M., & Maeir, A. (2022). Weekly calendar planning activity (WCPA): Validating a measure of functional cognition for adolescents with attention deficit hyperactivity disorder. *He American Journal of Occupational Therapy*, 76(6), 1–9.
- Franzen, M. D., & Wilhelm, K. L. (1996). Conceptual foundations of ecological validity in neuropsychological assessment. In R. J. Sbordone, & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 91–112). Gr Press/St Lucie Press, Inc, <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1996-98718-005&site=ehost-live>
- Gamito, P., Oliveira, J., Caires, C., Morais, D., Brito, R., Lopes, P., Saraiva, T., Soares, F., Sottomayor, C., Barata, F., Picareli, F., Prates, M., & Santos, C. (2015). Virtual kitchen test: Assessing frontal lobe functions in patients with alcohol dependence syndrome. *Methods in Information Medicine*, 2, 122–126.
- García-Molina, A., Tormos, J. M., Bernabeu, M., Junqué, C., & Roig-Rovira, T. (2012). Do traditional executive measures tell us anything about daily-life functioning after traumatic brain injury in Spanish-speaking individuals? *Brain Injury*, 26(6), 864–874. <https://doi.org/10.3109/02699052.2012.655362>
- Gass, C. S., Russell, E. W., & Hamilton, R. A. (1990). Accuracy of MMPI-based inferences regarding memory and concentration in closed-head-trauma patients. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 2(2), 175–178. <https://doi.org/10.1037/1040-3590.2.2.175>
- Gaylord-Ross, R. J. (1979). Mental retardation research, ecological validity, and the delivery of longitudinal education programs. *The Journal of Special Education*, 13(1), 69–80. <https://doi.org/10.1177/002246697901300112>
- Gilboa, Y., Jansari, A., Kerrouche, B., Uçak, E., Tiberghien, A., Benkhaled, O., Chevnard, M., & et al. (2019). Assessment of executive functions in children and adolescents with acquired brain injury (ABI) using a novel complex multi-tasking computerised task: The jansari assessment of executive functions for children (JEF-C©). *Neuropsychological Rehabilitation*, 29(9), 1359–1382. <https://doi.org/10.1080/09602011.2017.1411819>
- Gioia, G. A., & Isquith, P. K. (2004). Ecological assessment of executive function in traumatic brain injury. *Developmental Neuropsychology*, 25, 135–158. <https://doi.org/10.1080/87565641.2004.9651925>
- Hill, E. L., & Bird, C. M. (2006). Executive processes in asperger syndrome: Patterns of performance in a multiple case series. *Neuropsychologia*, 44(14), 2822–2835. <https://doi.org/10.1016/j.neuropsychologia.2006.06.007>
- Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The real-world approach and its problems: A critique of the term ecological validity. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.00721>
- Hoskin, K. M., Jackson, M., & Crowe, S. F. (2005). Can neuropsychological assessment predict capacity to manage personal finances? A comparison between brain impaired individuals with and without administrators. *Psychiatry, Psychology and Law*, 12(1), 56–67. <https://doi.org/10.1375/ppl.2005.12.1.56>
- Jansari, A. S., Devlin, A., Agnew, R., Akesson, K., Murphy, L., & Leadbetter, T. (2014). Ecological assessment of executive functions: A new virtual reality paradigm. *Brain Impairment*, 15(2), 71–85. <https://doi.org/10.1017/BrImp.2014.14>
- Jennings, J. W., & Keefer, L. H. (1969). Olfactory learning set in two varieties of domestic rat. *Psychological Reports*, 24(1), 3–15. <https://doi.org/10.2466/pr0.1969.24.1.3>
- Johnson, J. L. (1994). Episodic memory deficits in Alzheimer's disease: A behaviorally anchored scale. *Archives of Clinical Neuropsychology*, 9(4), 337–346. [https://doi.org/10.1016/0887-6177\(94\)90021-3](https://doi.org/10.1016/0887-6177(94)90021-3)
- Josman, N., Kizony, R., Hof, E., Goldenberg, K., Weiss, P. L., & Klinger, E. (2014). Using the virtual action planning-supermarket for evaluating executive functions in people with stroke. *Journal of Stroke and Cerebrovascular Disease*, 23(5), 879–887.
- Jovanovski, D., Zakzanis, K., Campbell, Z., Erb, S., & Nussbaum, D. (2012). Development of a novel, ecologically oriented virtual reality measure of executive function: The multitasking in the city test. *Applied Neuropsychology: Adult*, 19(3), 171–182. <https://doi.org/10.1080/09084282.2011.643955>
- Jovanovski, D., Zakzanis, K., Ruttan, L., Campbell, Z., Erb, S., & Nussbaum, D. (2012). Ecologically valid assessment of executive dysfunction using a novel virtual reality task in patients with acquired brain injury. *Applied Neuropsychology: Adult*, 19(3), 207–220. <https://doi.org/10.1080/09084282.2011.643956>
- Júlio, F., Ribeiro, M. J., Patrício, M., Malhão, A., Pedrosa, Fábio, Gonçalves, Hélio, Simões, M., van Asselen, M., Simões, Mário R., Castelo-Branco, M., & Januário, C. (2019). A novel ecological approach reveals early executive function impairments in Huntington's disease. *Frontiers in Psychology*, 10, Article 585. <https://doi.org/10.3389/fpsyg.2019.00585>
- Kallweit, C., Paucke, M., Strauß, M., & Exner, C. (2020). Cognitive deficits and psychosocial functioning in adult ADHD: Bridging the gap between objective test measures and subjective reports. *Journal of Clinical and Experimental Neuropsychology*, 42(6), 569–583. <https://doi.org/10.1080/13803395.2020.1779188>
- Kenworthy, L., Freeman, A., Ratto, A., Dudley, K., Powell, K. K., Pugliese, C. E., Anthony, L. G., & et al. (2020). Preliminary psychometrics for the executive function challenge task: A novel, “hot” flexibility, and planning task for youth. *Journal of the International Neuropsychological Society*, 26(7), 725–732. <https://doi.org/10.1017/S135561772000017X>
- Kibby, M. Y., Schmitter-Edgecombe, M., & Long, C. J. (1998). Ecological validity of neuropsychological tests: Focus on the California verbal learning test and the wisconsin card sorting test. *Archives of Clinical Neuropsychology*, 13(6), 523–534.
- Klinger, E., Chemin, I., Lebreton, S., & Marié, R. M. (2004). A virtual supermarket to assess cognitive planning. *Annual Review of CyberTherapy and Telemedicine*, 2, 49–57. <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2008-04557-007&site=ehost-live>
- Kourtesis, P., Collina, S., Doulmas, L. A. A., & MacPherson, S. E. (2021). Validation of the virtual reality everyday assessment lab (VR-EAL): An immersive virtual reality neuropsychological battery with enhanced ecological validity. *Journal of the International Neuropsychological Society*, 27(2), 181–196. <https://doi.org/10.1017/S1355617720000764>
- La Paglia, F., La Cascia, C., Rizzo, R., Cangialosi, F., Sanna, M., Riva, G., & La Barbera, D. (2014). Cognitive assessment of OCD patients: NeuroVR vs neuropsychological test. *Annual Review of CyberTherapy and Telemedicine*, 12, 40–44. <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2015-00374-007&site=ehost-live>

- La Paglia, F., La Cascia, C., Rizzo, R., Riva, G., & La Barbera, D. (2012). Assessment of executive functions in patients with obsessive compulsive disorder by neuroVR. *Annual Review of CyberTherapy and Telemedicine*, 10, 98–102. <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2014-01026-019&site=ehost-live>
- Laloux, J., Pellegrini, N., Mourad, H., Bertrand, H., Domken, M.-A., Van der Linden, M., & Larøi, F. (2013). Performance on a computerized shopping task significantly predicts real world functioning in persons diagnosed with bipolar disorder. *Psychiatry Research*, 210(2), 465–471. <https://doi.org/10.1016/j.psychres.2013.06.032>
- Laloux, J., Van der Linden, M., Levaux, M.-N., Mourad, H., Pirri, A., Bertrand, H., Domken, M.-A., Adam, S., & Larøi, F. (2014). Multitasking capacities in persons diagnosed with schizophrenia: A preliminary examination of their neurocognitive underpinnings and ability to predict real world functioning. *Psychiatry Research*, 217(3), 163–170.
- Lamberts, K. F., Evans, J. J., & Spikman, J. M. (2010). A real-life, ecologically valid test of executive functioning: The executive secretarial task. *Journal of Clinical and Experimental Neuropsychology*, 32(1), 56–65. <https://doi.org/10.1080/13803390902806550>
- Larrabee, G. J. (2015). The multiple validities of neuropsychological assessment. *The American Psychologist*, 70(8), 779–788. <https://doi.org/10.1037/a0039835>
- Latham, L. L. (1978). Construct and ecological validity of short-term memory measures in retarded persons. *American Journal of Mental Deficiency*, 83(2), 145–155.
- Lea, R. S., Benge, J. F., Adler, C. H., Beach, T. G., Belden, C. M., Zhang, N., Shill, H. A., Driver-Dunckley, E., Mehta, S. H., & Atri, A. (2021). An initial exploration of the convergent and ecological validity of the UDS 30 neuropsychological battery in parkinson's disease. *Journal of Clinical and Experimental Neuropsychology*, 43(9), 918–925. <https://doi.org/10.1080/13803395.2022.2034753>
- Logue, E., Marceaux, J., Balldin, V., & Hilsabeck, R. C. (2015). Further validation of the pillbox test in a mixed clinical sample. *The Clinical Neuropsychologist*, 29(5), 611–623. <https://doi.org/10.1080/13854046.2015.1061054>
- Long, C. J. (1996). Neuropsychological tests: A look at our past and the impact that ecological issues may have on our future. In R. J. Sbordone, & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 1–14). Gr Press/St Lucie Press, Inc. <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1996-98718-001&site=ehost-live>
- Longaud-Valès, A., Chevnard, M., Dufour, C., Grill, J., Puget, S., Sainte-Rose, C., Dellatolas, G., & et al. (2016). Assessment of executive functioning in children and young adults treated for frontal lobe tumours using ecologically valid tests. *Neuropsychological Rehabilitation*, 26(4), 558–583. <https://doi.org/10.1080/09602011.2015.1048253>
- Maeir, A., Krauss, S., & Katz, N. (2011). Ecological validity of the multiple errands test (MET) on discharge from neurorehabilitation hospital. *OTJR: Occupation, Participation, Health*, 31(1), S38–S46. <https://doi.org/10.3928/15394492-20101108-07>
- Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: Coming out of the office. *Brain Injury*, 18(11), 1067–1081. <https://doi.org/10.1080/02699050410001672387>
- Mitchell, M., & Miller, L. S. (2008). Prediction of functional status in older adults: The ecological validity of four Delis-Kaplan executive function system tests. *Journal of Clinical and Experimental Neuropsychology*, 30(6), 683–690. <https://doi.org/10.1080/13803390701679893>
- Montgomery, C., Hatton, N. P., Fisk, J. E., Ogden, R. S., & Jansari, A. (2010). Assessing the functional significance of ecstasy-related memory deficits using a virtual paradigm. *Human Psychopharmacology*, 25(4), 318–325. <https://doi.org/10.1002/hup.1119>
- Moriyama, Y., Mimura, M., Kato, M., Yoshino, A., Hara, T., Kashima, H., Kato, A., & Watanabe, A. (2002). Executive dysfunction and clinical outcome in chronic alcoholics. *Alcoholism: Clinical and Experimental Research*, 26(8), 1239–1244. <https://doi.org/10.1097/0000374-200208000-00016>
- Nadler Tzadok, Y., Eliav, R., Prtnoy, S., & Rand, D. (2022). Establishing the validity of the internet based bill-paying task to assess executive function deficits among adults with traumatic brain injury. *American Journal of Occupational Therapy*, 76(4), 1–10.
- Newcombe, F. (1987). Psychometric and behavioral evidence: Scope, limitations, and ecological validity. In H. S. Levin, J. Grafman, & H. M. Eisenberg (Eds.), *Neurobehavioral recovery from head injury* (pp. 129–145). Oxford University Press. <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1987-98716-009&site=ehost-live>
- Nir-Hadad, S. Y., Weiss, P. L., Waizman, A., Schwartz, N., & Kizony, R. (2017). A virtual shopping task for the assessment of executive functions: Validity for people with stroke. *Neuropsychological Rehabilitation*, 27(5), 808–833. <https://doi.org/10.1080/09602011.2015.1109523>
- Norris, G., & Tate, R. L. (2000). The behavioural assessment of the dysexecutive syndrome (BADS): Ecological, concurrent and construct validity. *Neuropsychological Rehabilitation*, 10(1), 33–45.
- O'Shea, R., Poz, R., Michael, A., Berrios, G. E., Evans, J. J., & Rubinsztein, J. S. (2010). Ecologically valid cognitive tests and everyday functioning in euthymic bipolar disorder patients. *Journal of Affective Disorders*, 125(1-3), 336–340. <https://doi.org/10.1016/j.jad.2009.12.012>
- Odhuba, R. A., Broek, M. D., & Johns, L. C. (2005). Ecological validity of measures of executive functioning. *British Journal of Clinical Psychology*, 44(2), 269–278. <https://doi.org/10.1348/014466505X29431>
- Oliveira, C. R., Lopes Filho, B., Jé, P., Sugarman, M. A., Esteves, C. S., Lima, M. M. B. M. P., Moret-Tatay, C., Irigaray, T. Q., & Argimon, I. I. L. (2016). Development and feasibility of a virtual reality task for the cognitive assessment of older adults: The ECO-VR. *The Spanish Journal of Psychology*, 19, Article E95. <https://doi.org/10.1017/sjp.2016.96>
- Orkin Simon, N., Jansari, A., & Gilboa, Y. (2022). Hebrew version of the jansari assessment of executive functions for children (JEF-C®): Translation, adaptation and validation. *Neuropsychological Rehabilitation*, 32(2), 287–305. <https://doi.org/10.1080/09602011.2020.1821718>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, Aørn, Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, 88, 105906. <https://doi.org/10.1016/j.ijsu.2021.105906>
- Parsons, T. D. (2011). Neuropsychological assessment using virtual environments: Enhanced assessment technology for improved ecological validity. In S. Brahmam (Ed.), *Advanced computational intelligence paradigms in healthcare: Virtual reality in psychotherapy, rehabilitation, and assessment* (pp. 271–289). Springer-Verlag.
- Parsons, T. D. (2015). Ecological validity in virtual reality-based neuropsychological assessment. In M. Khosrow-Pour (Ed.), *Information science and technology* (3rd ed., pp. 214–223). IGI Global.
- Pinto, J. O., Dores, A. R., Peixoto, B., & Barbosa, F. (2023). Ecological validity in neurocognitive assessment: Systematized review, content analysis, and proposal of an instrument. *Applied Neuropsychology*, online in advance of print. <https://doi.org/10.1080/23279095.2023.2170800>
- Pishdadian, S., Parlar, M. E., Heinrichs, R. W., & McDermid Vaz, S. (2022). An ecologically sensitive measure of executive cognition (the breakfast task) improves prediction of functional outcome in schizophrenia. *Applied Neuropsychology: Adult*, 29(5), 907–914. <https://doi.org/10.1080/23279095.2020.1821029>
- Poncet, F., Swaine, B., Taillefer, C., Lamoureux, J., Pradat-Diehl, P., & Chevnard, M. (2015). Reliability of the cooking task in adults with acquired brain injury. *Neuropsychological Rehabilitation*, 25(2), 298–317. <https://doi.org/10.1080/09602011.2014.971819>
- Portney, L. G., & Gross, K. D. (2020). Concept measurement validity. In L. G. Portney (Ed.), *Foundations of clinical research: Applications to evidence-based practice* (4th ed., pp. 127–140). Philadelphia, PA: F. A. Davis.
- Possin, K. L., LaMarre, A. K., Wood, K. A., Mungas, D. M., & Kramer, J. H. (2014). Ecological validity and neuroanatomical correlates of the NIH EXAMINER executive composite score. *Journal of the International Neuropsychological Society: JINS*, 20(1), 20–28. <https://doi.org/10.1017/S1355617713000611>
- Rabin, L. A., Burton, L. A., & Barr, W. B. (2007). Utilization rates of ecologically oriented instruments among clinical neuropsychologists. *The*

- Clinical Neuropsychologist*, 21(5), 727–743. <https://doi.org/10.1080/13854040600888776>
- Radomski, M. V., Davidson, L. F., Smith, L., Finkelstein, M., Cecchini, A., Heaton, K. J., McCulloch, K., Scherer, M., & Weightman, M. M. (2018). Toward return to duty decision-making after military mild traumatic brain injury: Preliminary validation of the charge of quarters duty test. *Military Medicine*, 183(7–8), e214–e222. <https://doi.org/10.1093/milmed/usx045>
- Rand, D., Rukan, S. B.-A., Weiss, P. L. (Tamar), & Katz, N. (2009). Validation of the virtual MET as an assessment tool for executive functions. *Neuropsychological Rehabilitation*, 19(4), 583–602. <https://doi.org/10.1080/09602010802469074>
- Raspelli, S., Pallavicini, F., Carelli, L., Morganti, F., Poletti, B., Corra, B., Silani, V., & G., Riva. (2011). Validation of a neuro virtual reality-based version of the multiple errands test for the assessment of executive functions. *Annual Review of CyberTherapy and Telemedicine*, 9, 72–76. <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2014-01025-017&site=ehost-live>
- Renison, B., Ponsford, J., Testa, R., Richardson, B., & Brownfield, K. (2012). The ecological and construct validity of a newly developed measure of executive function: The virtual library task. *Journal of the International Neuropsychological Society*, 18(3), 440–450. <https://doi.org/10.1017/S1355617711001883>
- Reynolds, B. W., Basso, M. R., Miller, A. K., Whiteside, D. M., & Combs, D. (2019). Executive function, impulsivity, and risky behaviors in young adults. *Neuropsychology*, 33(2), 212–221. <https://doi.org/10.1037/neu0000510>
- Roca, M., Parr, A., Thompson, R., Woolgar, A., Torralva, T., Antoun, N., Manes, F., & Duncan, J. (2010). Executive function and fluid intelligence after frontal lobe lesions. 133, 234–247. <https://doi.org/10.1093/brain/awp269>
- Roca, M., Torralva, T., Meli, F., Fiol, M., Calcagno, M.L., Carpintero, S., Pino, G.D., Vetrice, F., Martin, M.E., Vita, L., Correale, J. (2008). Cognitive deficits in multiple sclerosis correlate with changes in fronto-subcortical tracts. *Multiple Sclerosis*, 14, 364–369.
- Romundstad, B., Solem, S., Brandt, A. E., Hypher, R. E., Risnes, K., Rø, T. B., Stubberud, J., & Finnanger, T. G. (2022). Validity of the behavioural assessment of the dysexecutive syndrome for children (bads-c) in children and adolescents with pediatric acquired brain injury. *Neuropsychological Rehabilitation*, 33(4), 551–573. <https://doi.org/10.1080/09602011.2022.2034649>
- Rosenblum, S., Frisch, C., Deutsh-Castel, T., & Josman, N. (2015). Daily functioning profile of children with attention deficit hyperactive disorder: A pilot study using an ecological assessment. *Neuropsychological Rehabilitation*, 25(3), 402–418. <https://doi.org/10.1080/09602011.2014.940980>
- Rosetti, M. F., Ulloa, R. E., Reyes-Zamorano, E., Palacios-Cruz, L., de la Peña, F., & Hudson, R. (2018). A novel experimental paradigm to evaluate children and adolescents diagnosed with attention-deficit/hyperactivity disorder: Comparison with two standard neuropsychological methods. *Journal of Clinical and Experimental Neuropsychology*, 40(6), 576–585. <https://doi.org/10.1080/13803395.2017.1393501>
- Roy, A., Allain, P., Roulin, J.-L., Fournet, N., & Le Gall, D. (2015). Ecological approach of executive functions using the behavioural assessment of the dysexecutive syndrome for children (BADS-C): Developmental and validity study. *Journal of Clinical and Experimental Neuropsychology*, 37(9), 956–971. <https://doi.org/10.1080/13803395.2015.1072138>
- Salimpoor, V. N., & Desrocher, M. (2006). Increasing the utility of EF assessment of executive function in children. *Developmental Disabilities Bulletin*, 34(1-2), 15–42. <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2007-18602-003&site=ehost-live>
- Sanders, C., & Schmitter-Edgecombe, M. (2017). Examining the impact of formal planning on performance in older adults using a naturalistic task paradigm. *Neuropsychological Rehabilitation*, 27(5), 759–776. <https://doi.org/10.1080/09602011.2015.1107599>
- Sbordone, R. J. (1996). Ecological validity: Some critical issues for the neuropsychologist. In R. J. Sbordone, & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 15–41). Gr Press/St Lucie Press, Inc, <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1996-98718-002&site=ehost-live>
- Sbordone, R. J., & Long, C. J. (1996). *Ecological validity of neuropsychological testing*. Gr Press/St Lucie Press, Inc, <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1996-98718-000&site=ehost-live>
- Schmitter-Edgecombe, M., Cunningham, R., McAlister, C., Arrotta, K., & Weakley, A. (2021). The night out task and scoring application: An ill-structured, open-ended clinic-based test representing cognitive capacities used in everyday situations. *Archives of Clinical Neuropsychology*, 36(4), 537–553. <https://doi.org/10.1093/arclin/aca080>
- Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, 2(4), 419–436. [https://doi.org/10.1207/S15327078IN0204\\_02](https://doi.org/10.1207/S15327078IN0204_02)
- Scott, J. C., Woods, S. P., Vigil, O., Heaton, R. K., Schweinsburg, B. C., Ellis, R. J., Grant, I., Marcotte, T. D., & The San Diego HIV Neurobehavioral Research Center (HNRC) Group (2011). A neuropsychological investigation of multitasking in HIV infection: Implications for everyday functioning. *Neuropsychology*, 25(4), 511–519. <https://doi.org/10.1037/a0022491>
- Shallice, T., & Burgess, P.W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, 114(Pt 2), 727–741. <https://doi.org/10.1093/brain/114.2.727>
- Shimoni, M., Engel-Yeger, B., & Tirosh, E. (2012). Executive dysfunctions among boys with attention deficit hyperactivity disorder (ADHD): Performance-based test and parents report. *Research in Developmental Disabilities*, 33(3), 858–865. <https://doi.org/10.1016/j.ridd.2011.12.014>
- Siddiqui, I., Saperia, S., Fervaha, G., Da Silva, S., Jeffay, E., Zakzanis, K. K., Agid, O., Remington, G., & Foussias, G. (2019). Goal-directed planning and action impairments in schizophrenia evaluated in a virtual environment. *Schizophrenia Research*, 206, 400–406. <https://doi.org/10.1016/j.schres.2018.10.012>
- Silver, C.H. (2000). Ecological validity of neuropsychological assessment in childhood traumatic brain injury. *Journal of Head Traumatic Rehabilitation*, 15, 973–988. <https://doi.org/10.1097/00001199-200008000-00002>
- Silverberg, N. D., Hanks, R. A., & McKay, C. (2007). Cognitive estimation in traumatic brain injury. *Journal of the International Neuropsychological Society*, 13(5), 898–902. <https://doi.org/10.1017/S1355617707071135>
- Siu, A. F. Y., & Zhou, Y. (2014). Behavioral assessment of the dysexecutive syndrome for children: An examination of clinical utility for children with attention-deficit hyperactivity disorder (ADHD). *Journal of Child Neurology*, 29(5), 608–616. <https://doi.org/10.1177/0883073813516191>
- Spitoni, G. F., Aragonaa, M., Bevacqua, S., Cotugno, A., & Antonucci, G. (2018). An ecological approach to the behavioral assessment of executive functions in anorexia nervosa. *Psychiatry Research*, 259, 283–288. <https://doi.org/10.1016/j.psychres.2017.10.029>
- Spooner, D. M., & Pachana, N. A. (2006). Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations. *Archives of Clinical Neuropsychology*, 21(4), 327–337. <https://doi.org/10.1016/j.acn.2006.04.004>
- Steverson, T., Adlam, A. R., & Langdon, P. E. (2017). Development and validation of a modified multiple errands test for adults with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, 30(2), 255–268. <https://doi.org/10.1111/jar.12236>
- Suchy, Y. (2015). *Executive functions: A comprehensive guide for clinical practice*. Oxford University Press.
- Suchy, Y., Gereau Mora, M., DesRuisseaux, L. A., Niermeyer, M. A., & Lipio Brothers, S. (revision under review). *Pitfalls in research on ecological validity of novel executive function tests: A systematic review and a call to action*. Psychological Assessment.
- Suchy, Y., Gereau Mora, M., Lipio Brothers, S., & DesRuisseaux, L. A. (2024). Six elements test vs D-KEFS: What does “ecological validity” tell us? *Journal of the International Neuropsychological Society*. Online in advance of print.
- Suchy, Y., Lipio Brothers, S., DesRuisseaux, L. A., Gereau, M. M., Davis, J. R., Chilton, R. L. C., & Schmitter-Edgecombe, M. (2022). Ecological validity reconsidered: The night out task versus the D-KEFS. *Journal of Clinical and Experimental Neuropsychology*, 1-18(8), 562–579. <https://doi.org/10.1080/13803395.2022.2142527>
- Sudo, F. K., Alves, G. S., Ericeira-Valente, L., Alves, C. E. O., Tiel, C., Moreira, D. M., Laks, J., & Engelhardt, E. (2015). Executive testing predicts functional loss in subjects with white matter lesions. *Neurocase*, 21(6), 679–687. <https://doi.org/10.1080/13554794.2014.973884>

- Torralva, T., Gleichgerricht, E., Lischinsky, A., Roca, M., Manes, F. (2012). "Ecological" and highly demanding executive tasks detect real life deficits in high functioning adult ADHD patients. *Journal of Attentional Disorders*, 17(1), 1–10. <https://doi.org/10.1177/1087054710389988>
- Torralva, T., Gleichgerricht, E., Lischinsky, A., Roca, M., & Manes, F. (2013). Ecological" and highly demanding executive tasks detect real-life deficits in high-functioning adult ADHD patients. *Journal of Attention Disorders*, 17(1), 11–19. <https://doi.org/10.1177/1087054710389988>
- Torralva, T., Roca, M., Gleichgerricht, E., Bekinschtein, T., Manes, F. (2009). A neuropsychological battery to detect specific executive and social cognitive impairments in early frontotemporal dementia. *Brain*, 132, 1299–1309. <https://doi.org/10.1093/brain/awp041>
- Torralva, T., Strejilevich, S., Gleichgerricht, E., Roca, M., Martino, D., Cetrkovich, M., & Manes, F. (2012). Deficits in tasks of executive functioning that mimic real-life scenarios in bipolar disorder. *Bipolar Disorders*, 14(1), 118–125. <https://doi.org/10.1111/j.1399-5618.2012.00987.x>
- Toussaint-Thorin, M., Marchal, F., Benkhaled, O., Pradat-Diehl, P., Boyer, F. C., & Chevignard, M. (2013). Executive functions of children with developmental dyspraxia: Assessment combining neuropsychological and ecological tests. *Annals of Physical and Rehabilitation Medicine*, 56(4), 268–287.
- Tranel, D., Hathaway-Nepple, J., & Anderson, S. W. (2007). Impaired behavior on real-world tasks following damage to the ventromedial prefrontal cortex. *Journal of Clinical and Experimental Neuropsychology*, 29(3), 319–332. <https://doi.org/10.1080/13803390600701376>
- Tupper, D. E., & Cicerone, K. D. (1990). Introduction to the neuropsychology of everyday life. In D. E. Tupper & K. D. Cicerone (Eds.), *The neuropsychology of everyday life: Assessment and basic competencies*, (pp. 3–18). Kluwer Academic.
- Tyson, P. J., Laws, K. R., Flowers, K. A., Mortimer, A. M., & Schulz, J. (2008). Attention and executive function in people with schizophrenia: Relationship with social skills and quality of life. *International Journal of Psychiatry in Clinical Practice*, 12(2), 112–119. <https://doi.org/10.1080/13651500701687133>
- Valls-Serrano, C., Verdejo-García, A., Noël, X., & Caracul, A. (2018). Development of a contextualized version of the multiple errands test for people with substance dependence. *Journal of the International Neuropsychological Society*, 24(4), 347–359. <https://doi.org/10.1017/S1355617717001023>
- Van der Elst, W., Van Boxtel, M. P. J., Van Breukelen, G. J. P., & Jolles, J. (2008). A large-scale cross-sectional and longitudinal study into the ecological validity of neuropsychological test measures in neurologically intact people. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 23(7–8), 787–800. <https://doi.org/10.1016/j.acn.2008.09.002>
- Verdejo-García, A., & Pérez-García, M. (2007). Ecological assessment of executive functions in substance dependent individuals. *Drug and Alcohol Dependence*, 90(1), 48–55. <https://doi.org/10.1016/j.drugalcdep.2007.02.010>
- Ware, A. L., Crocker, N., O'Brien, J. W., Deweese, B. N., Roesch, S. C., Coles, C. D., Kable, J. A., May, P. A., Kalberg, W. O., Sowell, E. R., Jones, K. L., Riley, E. P., Mattson, S. N., & the CIFASD (2012). Executive function predicts adaptive behavior in children with histories of heavy prenatal alcohol exposure and attention-deficit/hyperactivity disorder. *Alcoholism, Clinical and Experimental Research*, 36(8), 1431–1441. <https://doi.org/10.1111/j.1530-0277.2011.01718.x>
- Webb, S. S., Jespersen, A., Chiu, E. G., Payne, F., Basting, R., Duta, M. D., & Demeyere, N. (2022). The oxford digital multiple errands test (OxMET): Validation of a simplified computer tablet based multiple errands test. *Neuropsychological Rehabilitation*, 32(6), 1007–1032. <https://doi.org/10.1080/09602011.2020.1862679>
- Werner, P., Rabinowitz, S., Klinger, E., Korczyn, A. D., & Josman, N. (2009). Use of the virtual action planning supermarket for the diagnosis of mild cognitive impairment: A preliminary study. *Dementia and Geriatric Cognitive Disorders*, 27(4), 301–309. <https://doi.org/10.1159/000204915>
- White, S. J., Burgess, P. W., & Hill, E. L. (2009). Impairments on "open-ended" executive function tests in autism. *Autism Research*, 2(3), 138–147. <https://doi.org/10.1002/aur.78>
- Wiedl, K. H., & Herrig, D. (1978). Ecological validity and scholastic success prognosis in learning and intelligence tests: A specimen study. *Diagnostica*, 24(2), 175–186. <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1979-32402-001&site=ehost-live>
- Wilson, B., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. J. (1996). *Behavioural assessment of the dysexecutive syndrome (BADS)*. Pearson.
- Wilson, B. A. (1993). Ecological validity of neuropsychological assessment: Do neuropsychological indexes predict performance in everyday activities? *Applied & Preventive Psychology*, 2(4), 209–215. [https://doi.org/10.1016/S0962-1849\(05\)80091-5](https://doi.org/10.1016/S0962-1849(05)80091-5)
- Wilson, B. A., Evans, J. J., Emslie, H., Alderman, N., & Burgess, P. (1998). The development of an ecologically valid test for assessing patients with dysexecutive syndrome. *Neuropsychological Rehabilitation*, 8(3), 213–228.
- Wolf, T. J., Dahl, A., Auen, C., & Doherty, M. (2017). The reliability and validity of the complex task performance assessment: A performance-based assessment of executive function. *Neuropsychological Rehabilitation*, 27(5), 707–721. <https://doi.org/10.1080/09602011.2015.1037771>
- Wood, R. L., & Liossi, C. (2006). The ecological validity of executive tests in a severely brain injured sample. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 21(5), 429–437. <https://doi.org/10.1016/j.acn.2005.06.014>
- Wood, R. L., & Liossi, C. (2007). The relationship between general intellectual ability and performance on ecologically valid executive tests in a severe brain injury sample. *Journal of the International Neuropsychological Society: JINS*, 13(1), 90–98. <https://doi.org/10.1017/S1355617707070129>
- Wood, R. L., & Bigler, E. (2017). Problems assessing executive dysfunction in neurobehavioural disability. In T. M. McMillan, & R. L. Wood (Eds.), *Neurobehavioural disability and social handicap following traumatic brain injury* (2nd ed. pp. 87–100). Routledge/Taylor & Francis Group, <https://doi.org/10.4324/9781315684710-7>
- Zartman, A. L., Hilsabeck, R. C., Guarnaccia, C. A., & Houtz, A. (2013). The pillbox test: An ecological measure of executive functioning and estimate of medication management abilities. *Archives of Clinical Neuropsychology*, 28(4), 307–319. <https://doi.org/10.1093/arclin/act014>
- Ziennik, R. E., & Suchy, Y. (2019). Ecological validity of performance-based measures of executive functions: Is face validity necessary for prediction of daily functioning? *Psychological Assessment*, 31(11), 1307–1318. <https://doi.org/10.1037/pas0000751>