



# Exploratory Factor Analysis of the NEPSY-II Conceptual Template: Acting on Evidence

Ryan J. McGill<sup>1</sup> · A. Alexander Beaujean<sup>2</sup> · Nicolas F. Benson<sup>3</sup> · Stefan C. Dombrowski<sup>4</sup> · Gary L. Canivez<sup>5</sup>

Received: 3 November 2023 / Revised: 28 January 2024 / Accepted: 5 February 2024  
© The Author(s) 2024

## Abstract

The present study examined the structure of the NEPSY-II within the norming sample using exploratory factor analysis. For the 3–4-year-old group, our results were conceptually uninterpretable. As a result, a unidimensional model was retained by default as a remedy to local fit issues. For the 7–12-year-old group, our analysis supported some aspects of the NEPSY-II conceptual domains in the form of a six-factor model that yielded the best fit to the data. While variance partitioning results indicate that the majority of NEPSY-II subtests at ages 7–12 contain adequate specificity to be interpreted in isolation, caution is suggested for interpreting the Social Perception subtests; in particular, given the inability to locate that latent dimension in either of the analyses conducted. Implications for the clinical interpretation of the instrument moving forward are discussed.

**Keywords** Structural validity · NEPSY-II · School neuropsychology · Applied validity

The second edition of the NEPSY (NEPSY-II; Korkman et al., 2007a) is an update of its predecessor (i.e., NEPSY-I; Korkman et al., 1998), which was one of the first individually-administered battery of tests designed to appraise neuropsychological development for children and adolescents.<sup>1</sup> The NEPSY authors state the instrument is useful for diagnostic decision-making and intervention planning for a variety of childhood disorders, such as Autism Spectrum Disorders, Specific Learning

Disorders, and Language Disorders (Korkman et al., 2007b). Although primarily created for clinicians with neuropsychology training, test authors state clinicians without such training can still employ the NEPSY-II to aid in clinical decision making (Kemp & Korkman, 2010). As such, the NEPSY-II portends to provide a variety of clinicians working with children a versatile array of tools to use in child assessment.

Both versions of the instrument were created based on Alexander Luria's (1973, 1980) theory of cognition and approach to clinical assessment. Very roughly, Luria held that the brain is comprised of various functional systems, which are interconnected neural regions that work together to support complex cognitive functions or processes. Since complex cognition requires the integration of neural structures and connections, cognitive difficulties will arise not only when there is trouble with integration, but also when there is trouble with the functioning of more basic structures. Thus, assessing a cognitive disorder requires separately assessing all the components that make up a functional system in order to isolate what is dysfunctional. On the NEPSY, the systems are referred to as domains, and subtests were created to assess basic components in the following domains: Attention and Executive Functioning, Language, Memory and Learning, Sensorimotor, and Visuospatial Processing. New to the NEPSY-II are subtests capturing components in the Social Perception domain.

<sup>1</sup> *NEPSY* is technically not an acronym, but an abbreviation that combines the five common letters in the word neuropsychology across multiple languages. We use the term *NEPSY-I* to refer to the first edition of the NEPSY, and use the unqualified *NEPSY* to refer to both editions generically.

✉ Ryan J. McGill  
rmcgill@wm.edu

<sup>1</sup> Department of School Psychology and Counselor Education, William & Mary, P. O. Box 8795, Williamsburg, VA, USA

<sup>2</sup> Department of Psychology, Baylor University, Waco, TX, USA

<sup>3</sup> Department of Educational Psychology, Baylor University, Waco, TX, USA

<sup>4</sup> Department of Graduate Education, Leadership, and Counseling, Rider University, Lawrenceville, USA

<sup>5</sup> Department of Psychology, Eastern Illinois University, Charleston, IL, USA

The NEPSY-II is a flexible battery, so the subtests administered to a given examinee will differ depending on kind of assessment the clinician chooses. The NEPSY-II allows for four different kinds of assessment. The first is a full assessment (i.e., using all the subtests available for a given age), which is supposed to provide a comprehensive neuropsychological evaluation and allow for identifying most consequences of brain pathology on a child's cognitive capacities. Since the full assessment is time and resource intensive, there are three abbreviated administration options (a) a general battery assessment that covers development in most of the neurocognitive domains; (b) one of the diagnostic battery assessments that involves administering a pre-selected set of subtests tailored to assesses specific referral concerns (e.g., reading disorder); or (c) a selective assessment featuring the administration of individual subtests chosen by the clinician. Irrespective of the assessment employed, scores from the administered subtests allow for creating client profiles of patterns of strengths and weaknesses across the domains (Matthews & Davis, 2018).

Specific revision goals for the NEPSY-II included improving domain coverage across the age span, enhancing the clinical and diagnostic utility, improving usability, and improving the overall psychometric properties (Korkman et al., 2007c). To accomplish this, four subtests from the NEPSY-I were eliminated and seven tests were added, for a total of 32 subtests.<sup>2</sup> In addition, all aggregate domain scores were dropped, which constrains interpretation to the "more clinically sensitive" subtest-level scores (Korkman et al., 2007c, p. 26).

## Structure of the NEPSY-II

Structure was a contentious issue for the NEPSY-I (e.g., Mosconi et al., 2008; Stinnett et al., 2002), and continues to be an issue for the NEPSY-II. First, theoretical justification for the development of the NEPSY domain structure is somewhat opaque. On the one hand, the domains appear to represent important areas of cognition and should be used to guide some aspects of score interpretation. On the other hand, the domains are downplayed by stating that a single subtest can be influenced by functions from several domains, and subtests within a single domain "may measure widely different abilities within the domain" (Korkman et al., 2007b, p. 80).

Second, evidence for structural claims, seemingly providing the empirical basis for the domains as an

organizational vehicle for the instrument is also inconsistent. Whereas it has been suggested that factor analysis is incompatible with the Lurian theory and assessment approach upon which the instrument is founded (Korkman, 1988), and no such analyses are provided to support the NEPSY structure. It is suggested that such analyses would be beneficial because it "would support data reduction strategies for research and would provide some sense of linkage to the theoretical model that underlies the NEPSY-II" (Kemp & Korkman, 2010, p. 239). In any case, provisional evidence for the NEPSY "structure" primarily consists of referencing Luria's theory or other neuropsychological findings, examining subtest content and response processes, and visual inspection of subtest correlation patterns of within a NEPSY domain and with scores from other psychological instruments (e.g., Wechsler Intelligence Scale for Children, Delis-Kaplan Executive Function System). While favoring visual inspection of correlations over multivariate methods is not unique (e.g., Beaujean & Parkin, 2022), it is not regarded as a compelling technique for elaborating on an instrument's potential underlying structure. For example, the 32 NEPSY-II subtests produce a matrix of 496 correlations, which is too much information for any human to comprehend well. Employing multivariate statistical methods can make the data more manageable and less prone to the spurious pitfalls of subjective inspection (Fabrigar & Wegener, 2012). This is why results from factor analyses and related techniques are typically provided as part of the portfolio of evidence supporting structural claims (Goodwin, 1999), and the *Joint Test Standards* strongly encourages furnishing this information in test Technical Manuals (American Educational Research Association et al., 2014).

Third, Luria's work was primarily with adults, so application with children should be seen as a hypothesis rather than fact, especially given rapid neurological and cognitive changes young children are undergoing (Karmiloff-Smith, 1998; see also Luria, 1973, Chapter 1). From that perspective, there are at least two competing hypotheses about NEPSY structure that come from the differentiation hypothesis and mutualism. The differentiation hypothesis states our intelligence spheres are relatively undifferentiated early in life, but become more specialized as we grow older and are exposed to various learning opportunities (Breit et al., 2022; Zimprich & Martin, 2010).<sup>3</sup> Thus, in younger children we would expect to find structure consisting of one (or just a few) relatively strong general attributes along with relatively weak group or specific attributes, but the specific/group attributes would become stronger during childhood and

<sup>2</sup> Some descriptions of the NEPSY-II state there are 36 subtests. This is because four memory subtests have delayed versions, so provides scores for both Immediate Recall and Delayed Recall.

<sup>3</sup> Technically, age and ability (knowledge) differentiations are two different hypotheses, but ability and age correlate strongly and positively in childhood, which confounds ability and age effects.

the general attributes would become weaker. By contrast, mutualism holds that our intelligence spheres initially have little-to-no structure, but structure emerges from repeatedly having certain kinds of experiences in which we employ certain cognitive processes together and there is mutually beneficial interactions among the cognitive processes (van der Maas et al., 2006). Thus, in younger children we would expect to find no general attributes, but one or more general attributes would emerge during childhood and grow stronger over time.

### Independent Factor Analytic Evidence for the NEPSY Instruments

A few structural investigations of the NEPSY instruments have been published, all of which employed some form of factor analysis. Stinnett and colleagues (2002) evaluated the structure of the NEPSY-I using exploratory factor analysis (EFA) of core subtest data from the norming sample ages 5–12. Extraction indices supported extracting 1–4 factors, but Stinnett et al. opted for the single-factor solution because all the multi-factor solutions had complexly-determined factors with multiple unanticipated cross-loadings among the indicators—both of which are common symptoms of over-extraction. They interpreted the single factor as reflecting an aggregate Language Comprehension dimension (loadings ranged from 0.26 to 0.64), and questioned whether the NEPSY authors' interpretation of subtests was viable due to lack of adequate reliable specific variance in the majority of those measures.

In an attempt to extend and replicate Stinnett and colleagues' work, Mosconi et al. (2008) evaluated the same norming data via confirmatory factor analysis (CFA) employing separate analyses for the total sample, a younger sample (ages 5–8) and an older cohort (ages 9–12). Like Stinnett et al., Mosconi et al. found no evidence supporting a five-factor model that would represent the original NEPSY domains, but unlike Stinnett et al., they found one-factor model fit very poorly (but see Dombrowski et al., 2021). Instead, they found a four-factor model fit the combined sample and younger sample well, but did not fit the older group well. Mosconi et al. concluded that the NEPSY-I structure differs by age and, at most, only four of the NEPSY conceptual domains are empirically identified.

Kervinen, (2015) recently examined the structure of Finnish version of the NEPSY-II using EFA in separate analyses of the 3–4-, 5–6-, and 7–15-year-old age groups in the norming sample. They found evidence supporting four factors in each of the age groups but the factors differed across the age span. There was a factor representing a Language domain in each age group, as well as a factor representing a combined Visuospatial/Sensorimotor domain. A third factor represented a mixture of Processing Speed and some other ability,

with the other ability differing at each age (i.e., fine motor control, working memory, fluency). The fourth factor varied substantially across the age groups. Interestingly, some subtests hung with the same subtests across age groups, while other subtests did not. Thus, not only did the structure differ across age groups, but also some of the subtests the factors were comprised.

### Purpose of Current Investigation

Given the unresolved questions pertaining to the integrity of the NEPSY theoretical model and the absence of any compelling evidence for structural claims about the NEPSY-II, it is necessary to clarify what the NEPSY-II scores capture and whether its conceptual template is a useful organizational framework for the battery. The NEPSY organizational framework of the instrument across six domains implies a theoretical structure supporting how clinicians should attribute some aspect of performance on the subtests in addition to informing decisions about which subtests to administer in various clinical situations (Tavakol & Wetzel, 2020). If subtests would fail to cohere with their assigned domain, it would raise fundamental questions as to what the NEPSY-II scores actually represent and whether the six-domain configurations is a viable organizing and interpretive framework for the instrument (Watkins, 2018). As such, evidence for structural claims is not optional because it subsequently impinges upon the theoretical template on which the instrument is based (Cattell, 1988).

More specifically, the hypotheses we investigate in this study are as follows.

- If the NEPSY domains are specified correctly, then we expect to see evidence for 6 factors in each dataset, and the factors should be interpretable along the lines of the 6 NEPSY-II domains (i.e., Attention and Executive Functioning, Language, Memory and Learning, Sensorimotor, Social Perception, and Visuospatial Processing).
- If the differentiation hypothesis is correct, then we expect to see evidence of one general factor in the younger age groups with a relatively strong factor loadings, and a general factor with weaker loadings in older age groups along with additional non-general factors.
- If the mutualism hypothesis is correct, then we expect to see evidence for a general factor with strong loadings in the older age groups, and in the younger age groups there should be evidence supporting either a general factor with weak loadings or non-general factors with strong loadings.

It is believed that the results provided by the present investigation will be instructive for furthering our

understanding how the measure should be interpreted and used in clinical practice.

## Method

### Participants

Participants were 1200 children aged 3–16 years who were included in the NEPSY-II American norming sample. This sample was obtained using a stratified sampling plan designed to accord with 2003 US Census estimates. Inspection of the demographic data reported in the Clinical Manual (Korkman et al., 2007c) reveal that the data for the norming sample was consistent with the US population parameters for age, sex, race/ethnicity, parent education level (as a proxy for socioeconomic status), and geographic region. There were 200–600 participants in each of the four age groups (ages 3–4, 5–6, 7–12, 13–16) that are the focus of the current study. For reasons that will be enumerated below, despite our intentions, the data for participants ages 5–6 and 13–16 were not able to be included in the present study.

### Instrument

The NEPSY-II (Korkman et al., 2007a, b, c) is a comprehensive neuropsychological assessment battery for children and adolescents ages 3–16. It contains 32 subtests apportioned to six functional domains (i.e., Attention and Executive Functioning, Language, Memory and Learning, Sensorimotor, Social Perception, Visuospatial Processing). Most subtests are comprised of multiple subcomponent skills, which allows for primary, process, contrast, and combined scores, and these scores can be expressed as scaled scores ( $M = 10$ ,  $SD = 3$ ), percentile ranks, or cumulative percentages. The scaled scores corresponding to the primary subtests in each age group were the focus of this investigation. The subtests available to administer differ by age (e.g., 18 subtests available for 3 years old, 21 subtests available for 16 years old). Organization and description of NEPSY-II scores across the six functional demands are outlined in Table 1.

Extensive norming and psychometric data can be found in the *NEPSY-II Clinical Manual* (Korkman et al., 2007c). Reliability evidence is mixed, and noticeably varies across age groups. For example, the subtests' stability coefficients (Appendix E, pp. 263–268) range from 0.21 to 0.91, despite having a relatively short retest interval (range = 12–51 days [ $M = 21$  days]). Likewise, only one subtest had an internal consistency estimate  $> 0.90$  in the 3–4-year-old group, six subtests met this criterion in the 5–6-year-olds, eight

subtests in the 7–12-year-olds, but only four subtests in the 13–16-year-olds. Overall, approximately only 80% of the estimates reported exceeding 0.70—which are considered marginal for a clinical instrument (Haynes et al., 2019). While heterogeneity in subtest reliability is not necessarily unusual, the subtest-level analyses that the NEPSY-II requires of users is dependent on these reliability indices; particularly the long-term stability of the obtained performance profiles (Russell et al., 2005; Styck et al., 2019).

It is noted that due to the length of the standardization battery, various NEPSY-II measures were not re-normed if (a) the subtests were unmodified from the NEPSY-I; and (b) no changes were expected in the norming as a result of the Flynn effect (Korkman et al., 2007c, pp. 38–39). Consequently, scores derived for some NEPSY-II subtests are based on norming data that was obtained over a quarter of a century ago. Although it is stated in the Technical Manual that most of the non-re-normed subtests are in the Sensorimotor domain, no information is provided about exactly what subtests are based on recycled 1998 NEPSY-I norms. Likewise, information is not provided about the procedures for how the data from two independent validation samples was successfully combined, or how potential missing data was treated in the total norming sample for the NEPSY-II. Regardless of the degree of rigor in which the NEPSY-II was evaluated for potential influence of the Flynn effect, such evaluation, alone, does not justify the decision to bypass developing current norms for all the subtests in the instrument (McGill et al., 2021). In any case, it is unknown what the effect of having mixed norming groups poses for the instrument.

### Procedure and Analyses

The NEPSY-II subtest scaled score data for four standardization age groups (ages 3–4, 5–6, 7–12, 13–16) were extracted from the intercorrelation matrices reported in the Clinical Manual (pp. 264–267, Tables E.1 to E.4) and subjected to exploratory factor analysis (EFA). We chose EFA for two reasons. First, it does not require a priori restricting any factor loadings to be zero, so allows subtests to cross-load on multiple factors (Manapat et al., 2023). This is consistent with the NEPSY authors discussion of the subtest–domain relations. Second, although we had a general idea of what the factor structure should look like under different developmental hypotheses, work on the NEPSY structure is too little and the results are too inconsistent for us to specify any strong structural hypotheses in advance. As such, our work here is more like that of a detective trying to establish a basis for future research (Behrens, 1997).

Consistent with best practices in EFA, we examined multiple criteria to determine the number of factors to retain, with additional consideration given to factor interpretability

**Table 1** Organization and implied structure of the NEPSY

Domain	Subtest/score	Abbreviation	Ages
Attention/Executive Functioning	Animal Sorting	AS	7–16
	Auditory Attention	AA	5–16
	Response Set	RS	7–16
	Clocks	CL	7–16
	Inhibition-Naming	INN	5–16
	Inhibition-Inhibition	INI	5–16
	Inhibition-Switching	INS	5–16
	Inhibition-Errors	INE	5–16
	Statue	ST	3–6
Language	Body Part Naming	BPN	3–4
	Body Part Identification	BPI	3–4
	Phonological Processing	PH	3–16
	Comprehension of Instructions	CI	3–16
	Speeded Naming	SN	3–16
Memory and Learning	Sentence Repetition	SR	3–6
	Narrative Memory	NM	3–16
	Word list Repetition	WRP	7–16
	Word list Recall	WRC	7–16
	Memory for Designs	MD	3–16
	Memory for Designs-Delayed	MDD	5–16
	Memory for Faces	MF	5–16
	Memory for Faces-Delayed	MFD	5–16
	Fingertip Tapping-Dominant	FTD	5–16
Sensorimotor	Fingertip Tapping-Nondominant	FTN	5–16
	Fingertip Tapping-Repetitions	FTR	5–16
	Fingertip Tapping-Sequences	FTS	5–16
	Visuomotor Precision	VP	3–12
	Affect Recognition	AR	3–16
Social Perception	Theory of Mind	TM	3–16
	Arrows	AW	5–16
Visuospatial Processing	Block Construction	BC	3–16
	Design Copying	DCP	3–16
	Geometric Puzzles	GP	3–16
	Picture Puzzles	PP	7–16

NEPSY-II alignment is based on the selective subtest/scores reported in the Technical Manual correlation matrices (Appendix E, pp. 263–268). Not all subtests/scores are included

as well as theoretical convergence in the resulting EFA solutions (Fabrigar et al., 1999; Sass & Schmitt, 2010). Specifically, we employed the visual scree test (Cattell, 1966), Horn's parallel analysis (HPA; Horn, 1965), minimum average partials (MAP; Velicer, 1976), Bayesian Information Criterion (BIC; Schwarz, 1978), and exploratory graph analysis (EGA; Golino & Epskamp, 2017).

Factor extraction tests were conducted using the *psych* (Revelle, 2023) and *EGAnet* (Golino et al., 2023) packages within the R Statistical System (R Core Team, 2023). As recommended by Keith et al. (2016), simulated eigenvalues for HPA were obtained using the principal axis factoring method. Next, principal axis EFA (Fabrigar, et al., 1999)

was used to analyze the NEPSY-II standardization sample correlation matrices using SPSS version 29 for Macintosh. Retained factors were subjected to promax rotation ( $k=4$ ; Gorsuch, 2003). Salient pattern loading coefficients were defined as those  $\geq 0.30$  (Child, 2006).

## Results

Results from Bartlett's Test of Sphericity (Bartlett, 1950) revealed that the correlation matrices for ages 3–4 ( $\chi^2 = 785.40$ ,  $df = 91$ ,  $p < 0.01$ ) and 7–12 ( $\chi^2 = 8472.61$ ,  $df = 406$ ,  $p < 0.01$ ) were not random. The

**Table 2** Number of factors suggested for extraction across different criteria by age group

Extraction criteria	NEPSY-II age group	
	3:0–4:11	7:0–12:11
Eigenvalue > 1	4	8
Scree test	4	8
Horn's parallel analysis	4	8
Minimum average partials	1	3
Exploratory graph analysis	4	6
Bayesian Information Criterion	2	8
Publisher implied	5–6*	6

Correlation matrices for ages 5:0–6:11 and 13:0–16:11 are non-positive definite and could not be analyzed

\*Attention/Executive Functioning only has one indicator at ages 3:0–4:11

Kaiser–Meyer–Olkin (KMO) statistic for those matrices were 0.673 and 0.840, respectively, both well above the minimum standard for conducting a factor analysis (Kaiser, 1974). Without standardization sample raw data, it was not possible to estimate skewness or kurtosis or determine if multivariate normality existed, but principal axis extraction does not assume normality which is preferable given the fact that the test authors indicate that some of the subtest score data is not normally distributed (Korkman et al., 2007c). Therefore, the correlation matrices for those age groups were deemed appropriate for the EFA procedures that were employed. Unfortunately, preliminary analyses of the intercorrelation matrices for ages 5–6 and 13–16 revealed that both matrices were non-positive definite and thus not able to be subjected to factor analysis in the present study (Lorenzo-Seva & Ferrando, 2021).

### Ages 3–4 EFA Analyses

Regarding the number of factors to extract, empirical criteria suggested four, two, and one factors as opposed to the five domains suggested by the test publisher (see Table 2 and Figure X.1<sup>4</sup>). While a presumed sixth factor is assumed conceptually by the battery configuration, that dimension would be produced from only a single indicator for the Attention and Executive Function domain which is mathematically impermissible in EFA. Wood et al. (1996) suggested that it is better to over-extract than under-extract, so EFA began by extracting five factors and then sequentially examined the adequacy of models with four, two, and one factor(s) after oblique rotation as suggested by empirical criteria (Table 3).

**Table 3** Ages 3–4 NEPSY-II general factor loadings

Subtest	Loading
Body Part Identification	.674
Body Part Naming	.640
Comprehension of Instructions	.636
Sentence Repetition	.614
Phonological Processing	.461
Theory of Mind	.540
Narrative Memory	.591
Block Construction	.467
Design Copying	.476
Memory for Designs	.439
Visuomotor Precision	.384
Affect Recognition	.633
Speeded Naming	.428
Statue	.367
Eigenvalue	4.7
Variance (%)	33.4

Un-rotated general structure coefficients assume a unidimensional model

Explication of all of the multidimensional models (see Table 4 and supplemental Tables X.1–2) resulted in symptoms of over-extraction in the form of fusion of theoretically meaningful constructs and salient cross-loadings on multiple dimensions, rendering the models unsatisfactory from an interpretive standpoint (Gorsuch, 2003). Among the competing models, the four-factor model (Table 4) was the only model to yield desired simple structure and was supported by the preponderance of the extraction criteria. Even so, interpretation of the factors was complicated by subtest score migration across the conceptual domains in virtually all of the factors that were extracted. As an example, Factor 2 was defined by multiple indicators from the Memory and Learning, Social Perception, and Visuospatial Processing domains with no discernable pattern in shared content or response processes across the scores. Even among the dimensions lending themselves to any coherent description, there was a merging of Visual-Motor tasks (Factor 4) and shared content in Body-Part Awareness and Control Indicators (i.e., Statue). It should be noted that the extracted communalities ranged from 0.200 (Statue) to 0.697 (Body Part Naming) and that such low values have been implicated as a source of instability in previous EFA analyses (e.g., Dombrowski et al., 2019).

As a result of these deficiencies, it was not possible to extrapolate any coherent linkages to the conceptual structure of the instrument from the multidimensional models that were examined and thus, only the unidimensional model was able to be retained as a matter of interpretive convenience like Stinnett et al. (2002). For the unidimensional model,

<sup>4</sup> Tables and figures denoted by an X refer to supplemental materials available at the following open source OSF link (<https://osf.io/wz3hu/>).

**Table 4** Ages 3–4 NEPSY-II principal axis factor with promax rotation (four factors)

Subtest	Factor				$h^2$	$u^2$	Error	$s^2$
	I	II	III	IV				
CI	<b>.769</b>	–.132	.096	.069	.613	.387	.140	.247
SR	<b>.576</b>	.182	–.046	.018	.453	.547	.110	.437
PH	<b>.571</b>	–.048	–.027	.082	.312	.688	.120	.568
SN	<b>.302</b>	.198	–.118	.171	.225	.775	.070	.705
NM	.144	<b>.691</b>	.000	–.206	.501	.499	.380	.119
MD	–.201	<b>.592</b>	.012	.179	.370	.630	.160	.470
AR	.110	<b>.493</b>	.118	.033	.432	.568	.250	.318
BC	–.034	<b>.493</b>	–.051	.211	.334	.666	.250	.416
TM	.227	<b>.384</b>	.075	–.082	.320	.680	.240	.440
BPN	–.049	.009	<b>.853</b>	.022	.697	.303	.190	.113
BPI	.178	–.016	<b>.732</b>	–.051	.690	.310	.270	.040
ST	–.169	.166	<b>.320</b>	.198	.200	.800	.180	.620
VP	.127	–.037	–.015	<b>.593</b>	.385	.615	.110	.505
DCP	.099	.050	.054	<b>.568</b>	.431	.569	.120	.449
Eigenvalue	4.7	1.4	1.2	1.0				
Variance (%)	33.4	9.8	8.3	7.3	42.5	57.4	18.5	38.9
Factor 1	1.0							
Factor 2	.56	1.0						
Factor 3	.62	.56	1.0					
Factor 4	.31	.46	.25	1.0				

$h^2$  = communality coefficient,  $u^2$  = uniqueness, error = 1 – reliability coefficient,  $s^2 = u^2$  – error. Pattern coefficients > .30 are bolded (Child, 2006). Structure coefficients omitted for clarity. *BPI* Body Part Identification, *BPN* Body Part Naming, *CI* Comprehension of Instructions, *SR* Sentence Repetition, *PH* Phonological Processing, *TM* Theory of Mind, *NM* Narrative Memory, *BC* Block Construction, *DCP* Design Copying, *MFD* Memory for Designs, *VP* Visuomotor Precision, *AR* Affect Recognition, *SN* Speeded Naming, *ST* Statue

only un-rotated  $g$ -loadings could be extracted ranging from 0.367 to 0.674 (see Table 1) which are poor based on Kaufman's (1994) criteria. The first eigenvalue accounted for the vast majority of total variance in the test (33.4%) and likely represents some undifferentiated ability dimension that is an artifact of explicating only a single "latent" dimension.

### Ages 7–12 First-order EFA Analyses

Empirical extraction criteria suggested three to eight factors with multiple criteria coalescing on six factors in accord with publisher theory (see Table 2 and Figure X.2). EFA proceeded by extracting eight factors and then sequentially examined the adequacy of competing models that were suggested by empirical extraction criteria. The three-factor model (Table X.3) resulted in desired simple structure and although a coherent Finger Tapping (i.e., Sensorimotor) dimension was recovered, the remaining two factors were complexly determined suggesting under-extraction. Conversely, the eight-factor model (Table X.4) appears to be over-extracted producing an eighth factor containing salient

bipolar loadings and a cross-loaded item (Finger Tapping-Repetition) which split from the aforementioned Finger Tapping dimension located previous models. Accordingly, the six-factor model was retained as the model best suited for explaining the data on the basis that it was supported by the majority of empirical criteria and yielded the only coherent conceptual alignment for the NEPSY-II among the rival models examined.

Table 5 presents results from extracting six NEPSY-II factors with promax ( $k=4$ ) rotation at ages 7–12. The extracted communalities ranged from 0.139 (affect recognition) to 0.910 (Finger Tapping-Dominant). Variance accounted for by the factors that were extracted ranged from 4.4 to 21.1% and correlations among those dimensions ranged from –0.04 to 0.60 indicating that a higher-order dimension is likely not tenable for these data. The extraction of six-factors produced desired simple structure, with minimal cross-domain migration, resulting in the following factors being identified: Factor 1 (Finger Tapping), Factor 2 (Visuospatial Processing), Factor 3 (Inhibition), Factor 4 (Memory/Language), Factor 5 (Design Memory), and Factor 6 (Facial Memory). Of note, animal sorting (Attention and Executive Function) migrated

**Table 5** Ages 7–12 NEPSY-II principal axis factor with promax rotation (six factors)

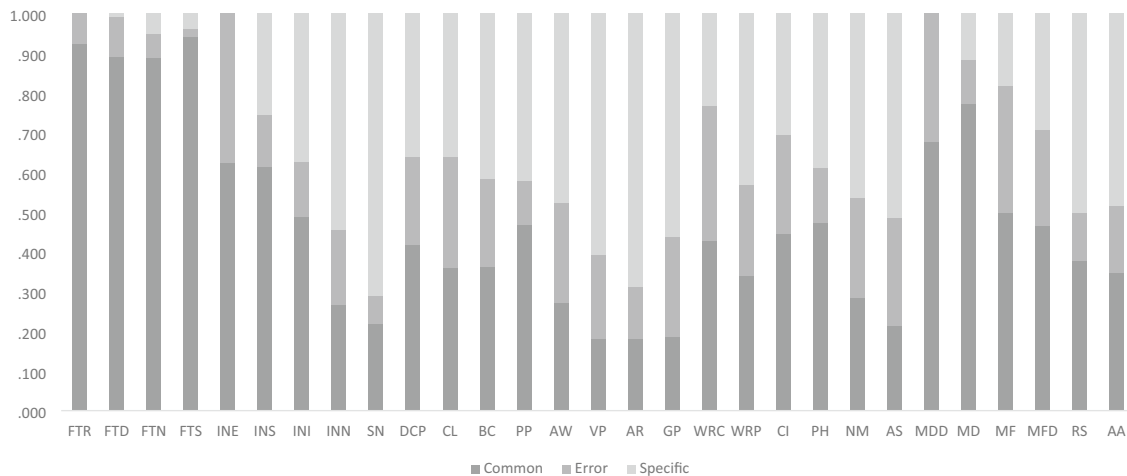
Subtest	Factor						$h^2$	$u^2$	Error	$s^2$
	I	II	III	IV	V	VI				
FTD	<b>.961</b>	.027	–.015	–.045	–.010	–.004	.910	.090	.100	.000
FTN	<b>.947</b>	–.031	.015	.023	.018	.007	.907	.093	.060	.033
FTR	<b>.868</b>	.126	–.058	–.142	–.090	–.001	.740	.260	.080	.180
FTS	<b>.800</b>	–.124	.065	.146	.077	–.005	.708	.292	.020	.272
DCP	.052	<b>.641</b>	–.051	–.084	–.007	.068	.353	.647	.220	.427
BC	.080	<b>.633</b>	–.096	–.020	.078	–.125	.345	.655	.220	.435
PP	–.042	<b>.549</b>	.102	.018	.010	.147	.470	.530	.110	.420
AW	–.083	<b>.530</b>	–.033	.007	–.034	.039	.273	.727	.250	.477
CL	.039	<b>.486</b>	.140	.058	–.050	–.067	.333	.667	.280	.387
GP	–.035	<b>.399</b>	–.011	.061	–.016	.014	.185	.815	.250	.565
AS	.017	<b>.302</b>	–.014	.224	–.007	–.044	.202	.798	.270	.528
VP	–.047	.287	.102	–.038	.021	–.045	.110	.890	.210	.680
AR	.048	.222	.045	.045	.039	.135	.139	.861	.130	.731
INE	–.087	–.018	<b>.983</b>	–.113	–.008	–.024	.843	.157	.520	.000
INS	–.004	.041	<b>.740</b>	.069	–.044	–.043	.614	.386	.130	.256
INI	.035	.109	<b>.621</b>	–.069	.022	.037	.444	.556	.140	.416
INN	.009	–.066	<b>.528</b>	.021	.031	.000	.265	.735	.190	.545
RS	.130	–.078	<b>.318</b>	.208	.020	.004	.209	.791	.120	.671
SN	.083	–.010	.263	.262	–.017	.017	.217	.783	.070	.713
WRP	.000	–.135	–.034	<b>.674</b>	–.024	.005	.341	.659	.230	.429
WRC	–.072	.032	–.091	<b>.634</b>	.003	–.004	.365	.635	.340	.295
CI	–.009	.139	–.019	<b>.604</b>	–.045	–.004	.446	.554	.250	.304
PH	–.023	.191	.042	<b>.499</b>	.079	–.016	.471	.529	.140	.389
NM	–.026	.143	.022	<b>.438</b>	–.012	–.085	.272	.728	.250	.478
AA	.039	–.083	.109	<b>.423</b>	–.008	.131	.240	.760	.170	.590
MD	–.019	.063	.000	–.029	<b>.867</b>	–.032	.773	.227	.110	.117
MDD	.004	–.031	.005	–.006	<b>.863</b>	.037	.739	.261	.360	.000
MF	–.010	–.033	–.067	.072	–.027	<b>.709</b>	.500	.500	.320	.180
MFD	.005	.049	.037	–.062	.034	<b>.659</b>	.458	.542	.240	.302
Eigenvalue	6.1	3.5	1.9	1.5	1.4	1.3				
Variance (%)	21.1	11.9	6.7	5.2	4.7	4.4	44.4	55.6	19.9	37.3
Factor 1	1.0									
Factor 2	.05	1.0								
Factor 3	.08	.5	1.0							
Factor 4	.17	.60	.50	1.0						
Factor 5	–.04	.46	.28	.35	1.0					
Factor 6	.02	.32	.10	.26	.30	1.0				

$h^2$ =communality coefficient,  $u^2$ =uniqueness, error=1-reliability coefficient,  $s^2=u^2$ -error. Pattern coefficients > .30 are bolded (Child, 2006). Structure coefficients omitted for clarity. *FTR* Finger Tapping-Repetition, *FTN* Finger Tapping-Dominant, *FTS* Finger Tapping-Sequences, *INE* Inhibition-Errors, *INS* Inhibition-Switching, *INI* Inhibition-Inhibition, *INN* Inhibition-Naming, *SN* Speeded Naming, *DCP* Design Copying, *CL* Clocks, *BC* Block Construction, *PP* Picture Puzzles, *AW* Arrows, *VP* Visuomotor Precision, *AR* Affect Recognition, *GP* Geometric Puzzles, *WRC* Word List Recall, *WRP* Word List Repetition, *CI* Comprehension of Instructions, *PH* Phonological Processing, *NM* Narrative Memory, *AS* Animal Sorting, *MDD* Memory for Designs-Delayed, *MD* Memory for Designs, *MF* Memory for Faces, *MFD* Memory for Faces-Delayed, *RS* Response Set, *AA* Auditory Attention

to Factor 3 as a lone outlier and affect recognition failed to load saliently on any latent dimension. These results provide some support for aspects of alignment with the hypothesized

NEPSY-II model but it is likely that Factors 1 and 3 represent method factors as opposed to a broader psychological dimension (e.g., Sensorimotor).

Sources of Variance in NEPSY-II Subtests (ages 7:0-12:11)



**Fig. 1** Sources of variance in NEPSY-II subtests (ages 7:0–12:11). *Note.* FTR, Finger Tapping-Repetition; FTD, Finger Tapping-Dominant; FTN, Finger Tapping-Nondominant; FTS, Finger Tapping-Sequences; INE, Inhibition-Errors; INS, Inhibition-Switching; INI, Inhibition-Inhibition; INN, Inhibition-Naming; SN, Speeded Naming; DCP, Design Copying; CL, Clocks; BC, Block Construction; PP, Picture Puzzles; AW, Arrows; VP, Visuomotor Precision; AR, Affect Recognition; GP,

Geometric Puzzles; WRC, Word List Recall; WRP, Word List Repetition; CI, Comprehension Of Instructions; PH, Phonological Processing; NM, Narrative Memory; AS, Animal Sorting; MDD, Memory for Designs-Delayed; MD, Memory for Designs; MF, Memory for Faces; MFD, Memory for Faces-Delayed; RS, Response Set; AA, Auditory Attention

## Variance Partitioning Results

The use of EFA permits the assignment of variance at different levels of generality (i.e., group-specific factors and subtest-level). Furnishing subtest specificity estimates (i.e., the component of reliable test performance that is unique to that test after the error term in the measure is extracted) are particularly instructive for determining which tests are suitable for the Lurian interpretive procedures described in the Clinical Manual (Korkman et al., 2007c). For the six-factor solution at ages 7–12, the common factors absorbed 44.4% of the total variance in the measure. Conversely, the subtests contained larger portions of reliable variance that were unique to the scores. Kaufman (1994) has proposed that uniqueness may be considered *high* when an individual test's unique variance was equal to or above 25% of the total variance for the test, and that component exceeds the subtest's corresponding error variance. As can be seen graphically in Fig. 1, a vast majority of NEPSY-II scores meet or exceed this criterion at ages 7–12.

## Discussion

The present study examined the internal structure of the NEPSY-II subtest scores for participants in the standardization normative sample (ages 3–16) using EFA procedures. This is the first published structural validity investigation of

the NEPSY-II since its publication as no direct examination of internal structure is reported in the test's Clinical Manual (Korkman et al., 2007c). Instead, users must extrapolate what the NEPSY-II measures from the conceptual organization of the test battery and descriptive information provided in NEPSY-II interpretive materials (e.g., Matthews & Davis, 2018). Although the conceptual organization of the test battery implies that it measures a diverse array of neurocognitive functions across six functional domains, previous independent factor analytic investigations of the NEPSY yielded conflicting findings with respect to its psychological dimensionality (e.g., Mosconi et al., 2008; Stinnett et al., 2002). As a consequence, questions remain as to what the instrument actually measures and how it should be interpreted and used in clinical settings. Given that it has been argued that that users of the NEPSY-II would benefit from a comprehensive factor analysis (Matthews et al., 2012), the present investigation was designed, in part, to help fill that gap in the literature.

Results from the present study largely comport with those furnished by previous investigations suggesting that the NEPSY-II measurement model is likely not invariant and at times lacks coherence with the conceptual template outlined in the Clinical Manual across the age span. At ages 3–4, empirical extraction criteria failed to align with the five domains posited by the test publisher. Attempts to examine rival multidimensional structures was complicated by lack of theoretical linkage for the results, real or implied, by the

NEPSY-II organizing template resulting in the retention of a single-factor model for the data. Whereas Stinnett and colleagues (2002) produced similar results for the entire standardization sample for the NEPSY, concluding that the instrument was dominated by a single Language Comprehension dimension, the results of this study suggest that the nature of that construct is less well understood given the weak subtest loadings on that dimension (Larson et al., 1988). Although Korkman et al. (2007c) suggest that Language plays a dominant role in performance for many NEPSY-II tasks, it is likely not singularly sufficient for explaining performance at ages 3–4 (Tideman & Gustafsson, 2004). Nevertheless, the weak general factor at ages 3–4 and the lack of compelling evidence for a strong general factor underlies the data at ages 7–12, suggests that both the mutualism and differentiation hypotheses do not provide an adequate explanation for the development of NEPSY-II abilities across the age span.

At ages 7–12, the results cohere in part with those furnished by Mosconi et al. (2008) using CFA on the NEPSY in targeted analyses of the normative sample. Whereas their analyses suggested that a single-factor model was potentially tenable for the data, the present EFA results did not yield compelling evidence for a single-general factor at that age given the presence of mostly weak correlations in the best fitting oblique factors solution. EFA analyses provided more compelling evidence for alignment of NEPSY-II measures with their conceptual organization with respect to the Language, Sensorimotor (i.e., Finger Tapping), Visuospatial Processing, Attention and Executive Function, and Memory and Learning domains at those ages. However, no compelling evidence was found to support the posited alignment of the Social Perception measures in either age group suggesting caution in their use and interpretation as indicators of social-emotional functioning (e.g., speculation about Autism). It should also be noted that Memory for Faces and Memory for Designs formed distinct factors after aligning with their related delayed tasks. Similar “splitting” has also been observed in EFA studies of other ability measures sampling Memory and Learning domains (e.g., McGill & Dombrowski, 2018). Additional research is needed to determine if those recovered factors reflect viable psychological constructs or are merely an artifact of shared content between the measures similar to the Finger Tapping tasks.

As the NEPSY-II no longer yields domain-based scores and users are encouraged to interpret the instrument primary at the subtest-level, the robust average specificity component at ages 7–12 (37.3%) lends some psychometric support to that approach, indicating that most of the NEPSY-II scores are likely best interpreted on the basis of their task-specific elements rather than alignment with broader latent constructs. Even so, given that the NEPSY-II is commonly utilized in school-based settings by practitioners (i.e., school psychologists) who likely do not have the necessary

advanced training in the intricacies of *true* neuropsychological assessment which the Lurian interpretive approach requires (Jantz & Plotts, 2014), it introduces questions about the inherent risk of some users over-interpreting results from the measure to support clinical hypotheses. This concern becomes particularly salient when targeted measures are administered selectively in a cross-battery assessment framework (e.g., Flanagan et al., 2013) and NEPSY-II test results fail to cohere with other test data. To wit, Korkman et al. (2007a, b, c) stress that an observed weakness on any particular NEPSY-II subtest should not be interpreted as de facto evidence of a deficit in the broader domain thought to be sampled by the test. This perspective is supported by base rate evidence furnished by Brooks et al. (2010) who found that low NEPSY-II scores are common in otherwise healthy children suggesting that some of the measures may produce an unacceptable number of false positive test results. In addition, users must also take into consideration that omnibus reliability coefficients for many of the NEPSY-II subtest scores remain unacceptable for clinical interpretation regardless of the level of specificity contained in the tests.

## Limitations

As with any study, several limitations must be considered when interpreting the present results. First, it remains unclear as to why the matrices corresponding to the normative data for participants ages 5–6 and 13–16 failed to converge in EFA. As singularity is often the default source when a matrix is found to be non-positive definite and unable to be inverted (Lorenzo-Seva & Ferrando, 2021), we again inspected the values in the matrices though were unable to identify any coefficients that appear to be offending. It is worth noting that inspection of the normative conversion tables in the Clinical Manual (Korkman et al., 2007a, b, c) reveal inadequate item density and ceiling effects in several of the NEPSY-II measures at ages 13–16. Thus, it is possible that this restricted scaling may have led to the collapse in latent dimensionality in the measure at that age. It is also worth disclosing that the decision to extend the age range for the test was made post-hoc after preliminary pilot studies were completed and does not appear to be part of the original design plan. Nevertheless, due to the inability to examine the factor structure of the instrument at those ages, it remains unclear as to whether the differences in dimensional complexity observed across the age range are due to developmental differences in the growth and emergence of higher-order abilities (as postulated by Tucker-Drob, 2009) or an artifact of differential measurement-effects across the age range.

Additionally, during preliminary validation of the NEPSY-I, Korkman (1988) contended that factor analysis was ill-suited to uncover the true latent structure of the instrument.

However, Stinnett et al. (2002) counter that despite references to the complexities of the Lurian model, it is unclear how that model is represented conceptually in the way the battery was organized and designed. Further, there is nothing particularly complicated about the template and organization of the domains that would ostensibly render them unrecoverable in a factor analytic investigation and omnibus theory would suggest that the domains are organized in a way conceptually as to promote moderate to high correlations among the indicators that are more than amenable to factor analysis (e.g., Carroll, 1993). Nevertheless, given the questions raised in this study and in previous NEPSY research as to what the battery actually measures, future research on the measure using emerging psychometric methods able to capture un-modeled complexity in conventional factor analytic investigations such as network analysis (i.e., Borsboom, 2022) would be a welcome addition to the literature on the matter. Particularly at ages 3–4 where the underlying dimensional structure of the test lacks clarity.

Finally, although the Finger Tapping factor observed at ages 7–12 was well defined demonstrating the strongest subtest alignment of any of the models examined, use of those measures in clinical practice is likely contraindicated by the lack of current normative data on which to anchor test performance for those indicators. As previously mentioned, the data upon which those measures were developed is now well over 25 years old and thus it is unclear the degree to which those data comport with current reference samples on related measures. In spite of the argumentation contained in the Clinical Manual, one cannot simply extrapolate current from past performance with that large of a time lag in the life-cycle of a commercial ability measure that is used to render current diagnostic decisions.

## Conclusion

In sum, the authors of the NEPSY-II should be lauded for addressing a number of criticisms of the NEPSY in the revised version of the instrument. Most notably, domain-scores have been dropped in response to questions raised about the dimensional complexity of the instrument and specificity in the subtests at ages 7–12 is more than adequate to support the interpretive-focus stressed in the Clinical Manual at that level of the test. Nevertheless, the present EFA results suggest that, at a minimum, the latent structure of the measure is likely not invariant across age and although there is some evidence that at ages 7–12, the measure appears to comport align in part with its posited conceptual template, caution is urged in administering and interpreting measures from the hypothesized Sensorimotor and Social Processing domains.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40817-024-00158-8>.

**Funding** This research was partially supported by a faculty research grant from William & Mary.

## Declarations

**Conflict of Interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- American Educational Research Association, American Psychological Association, & National Council on Measurement on Education. (2014). Standards for educational and psychological testing. *American Educational Research Association*.
- Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, 3(2), 77–85. <https://doi.org/10.1111/j.2044-8317.1950.tb00285.x>
- Beaujean, A. A., & Parkin, J. R. (2022). Evaluation of the Wechsler individual achievement test-fourth edition as a measurement instrument. *Journal of Intelligence*, 10(2). <https://doi.org/10.3390/jintelligence10020030>
- Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, 2(2), 131–160. <https://doi.org/10.1037/1082-989X.2.2.131>
- Borsboom, D. (2022). Possible futures for network psychometrics. *Psychometrika*, 87(1), 253–265. <https://doi.org/10.1007/s11336-022-09851-z>
- Breit, M., Brunner, M., Molenaar, D., & Preckel, F. (2022). Differentiation hypotheses of intelligence: A systematic review of the empirical evidence and an agenda for future research. *Psychological Bulletin*, 148(7–8), 518–554. <https://doi.org/10.1037/bul0000379>
- Brooks, B. L., Sherman, E. M. S., & Iverson, G. L. (2010). Healthy children get low scores too: Prevalence of low scores on the NEPSY-II in preschoolers, children, and adolescents. *Archives of Clinical Neuropsychology*, 25(3), 182–190. <https://doi.org/10.1093/arclin/acq005>
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511571312>
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2), 245–276. [https://doi.org/10.1207/s15327906mbr0102\\_10](https://doi.org/10.1207/s15327906mbr0102_10)
- Cattell, R. B. (1988). The meaning and strategic use of factor analysis. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 131–203). Plenum Press. [https://doi.org/10.1007/978-1-4613-0893-5\\_4](https://doi.org/10.1007/978-1-4613-0893-5_4)
- Child, D. (2006). *Essentials of factor analysis* (3<sup>rd</sup> ed.). Continuum.

- Dombrowski, S. C., Beaujean, A. A., McGill, R. J., Benson, N. F., & Schneider, W. J. (2019). Using exploratory bifactor analysis to understand the latent structure of multidimensional psychological measures: An example featuring the WISC-V. *Structural Equation Modeling*, 26(6), 847–869. <https://doi.org/10.1080/10705511.2019.1622421>
- Dombrowski, S. C., McGill, R. J., Canivez, G. L., Watkins, M. W., & Beaujean, A. A. (2021). Factor analysis and variance partitioning in intelligence research: Clarifying misconceptions. *Journal of Psychoeducational Assessment*, 39(1), 28–38. <https://doi.org/10.1177/0734282920961952>
- Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford University Press.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299. <https://doi.org/10.1037/1082-989X.4.3.272>
- Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). John Wiley.
- Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS ONE*, 12(6), Article e0174035. <https://doi.org/10.1371/journal.pone.0174035>
- Golino, H., Christensen, A. P., Moulder, R., Garrido, L. E., Jamison, L., Shi, D. (2023). EGAnet: Exploratory graph analysis-A framework for estimating the number of dimensions in multivariate data using network psychometrics [Version 2.0.1]. R Foundation for Statistical Computing.
- Goodwin, L. D. (1999). The role of factor analysis in the estimation of construct validity. *Measurement in Physical Education and Exercise Science*, 3(2), 85–100. [https://doi.org/10.1207/s15327841mpee0302\\_2](https://doi.org/10.1207/s15327841mpee0302_2)
- Gorsuch, R. L. (2003). Factor analysis. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (Vol. 2, pp. 143–164). John Wiley.
- Haynes, S. N., Smith, G. T., & Hunsley, J. D. (2019). *Scientific foundations of clinical assessment* (2nd ed.). Routledge.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. <https://doi.org/10.1007/BF02289447>
- Jantz, P. B., & Plotts, C. A. (2014). Integrating neuropsychology and school psychology: Potential and pitfalls. *Contemporary School Psychology*, 18(1), 69–80. <https://doi.org/10.1007/s40688-013-0006-2>
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36. <https://doi.org/10.1007/BF02291575>
- Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*, 2(10), 389–398. [https://doi.org/10.1016/S1364-6613\(98\)01230-3](https://doi.org/10.1016/S1364-6613(98)01230-3)
- Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. John Wiley.
- Keith, T. Z., Caemmerer, J. M., & Reynolds, M. R. (2016). Comparison of methods for factor extraction for cognitive test-like data: Which overfactor, which underfactor? *Intelligence*, 54, 37–54. <https://doi.org/10.1016/j.intell.2015.11.003>
- Kemp, S. L., & Korkman, M. (2010). *Essentials of NEPSY-II assessment*. Wiley.
- Kervinen, S. (2015). Exploration of the internal structure of the NEPSY-II [Master's thesis, University of Helsinki].
- Korkman, M. (1988). NEPSY: An adaptation of Luria's investigation for young children. *Clinical Neuropsychologist*, 2(4), 375–392. <https://doi.org/10.1080/13854048808403275>
- Korkman, M., Kirk, U., & Kemp, S. (1998). NEPSY: A developmental neuropsychological assessment. *Psychological Corporation*.
- Korkman, M., Kirk, U., & Kemp, S. (2007a). NEPSY-Second Edition. NCS Pearson.
- Korkman, M., Kirk, U., & Kemp, S. (2007b). NEPSY-Second Edition administration manual. NCS Pearson.
- Korkman, M., Kirk, U., & Kemp, S. (2007c). NEPSY-Second Edition clinical and interpretive manual. NCS Pearson.
- Larson, G. E., Merritt, C. R., & Williams, S. E. (1988). Information processing and intelligence: Some implications of task complexity. *Intelligence*, 12(2), 131–147. [https://doi.org/10.1016/0160-2896\(88\)90012-8](https://doi.org/10.1016/0160-2896(88)90012-8)
- Lorenzo-Seva, U., & Ferrando, P. J. (2021). Not positive definite correlation matrices in exploratory item factor analysis: Causes, consequences and a proposed solution. *Structural Equation Modeling*, 28(1), 138–147. <https://doi.org/10.1080/10705511.2020.1735393>
- Luria, A. R. (1973). *The working brain: An introduction to neuropsychology*. Penguin Books.
- Luria, A. R. (1980). *Higher cortical functions in man* (2nd ed.). Basic Books.
- Manapat, P. D., Anderson, S. F., & Edwards, M. C. (2023). *Evaluating avoidable heterogeneity in exploratory factor analysis results*. Advance online publication. <https://doi.org/10.1037/met0000589>
- Matthews, R. N., & Davis, J. L. (2018). The NEPSY-II. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 553–566). Guilford Press.
- Matthews, R. N., Riccio, C. A., & Davis, J. L. (2012). The NEPSY-II. In D. P. Flanagan & Harrison, P. L. (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 422–435). Guilford Press.
- McGill, R. J., & Dombrowski, S. C. (2018). Factor structure of the CHC model for the KABC-II: Exploratory factor analyses with the 16 core and supplemental subtests. *Contemporary School Psychology*, 22(3), 279–293. <https://doi.org/10.1007/s40688-017-0152-z>
- McGill, R. J., Ward, T. J., & Canivez, G. L. (2021). On the validation of ability measures in school psychology: Do established psychometric standards matter? *School Psychology International*, 42(2), 210–216. <https://doi.org/10.1177/0143034320985209>
- Mosconi, M., Nelson, L., & Hooper, S. R. (2008). Confirmatory factor analysis of the NEPSY for younger and older school-age children. *Psychological Reports*, 102(3), 861–866. <https://doi.org/10.2466/pr0.102.3.861-866>
- R Core Team. (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Retrieved from <https://www.Rproject.org>
- Revelle, W. (2023). psych: Procedures for psychological, psychometric, and personality research. [Version 2.3.9]. R Foundation for Statistical Computing.
- Russell, E. W., Russell, S. L. K., & Hill, B. D. (2005). The fundamental psychometric status of neuropsychological batteries. *Archives of Clinical Neuropsychology*, 20(6), 785–794. <https://doi.org/10.1016/j.acn.2005.05.001>
- Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research*, 45(1), 73–103. <https://doi.org/10.1080/00273170903504810>
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Stinnett, T. A., Oehler-Stinnett, J., Fuqua, D. R., & Palmer, L. S. (2002). Examination of the underlying structure of the NEPSY: A developmental neuropsychological assessment. *Journal of Psychoeducational Assessment*, 20(1), 66–82. <https://doi.org/10.1177/073428290202000105>
- Styck, K. M., Beaujean, A. A., & Watkins, M. W. (2019). Profile reliability of cognitive ability subscores in a referred sample. *Archives of Scientific Psychology*, 7(1), 119–128. <https://doi.org/10.1037/arc0000064>

- Tavakol, M., & Wetzel, A. (2020). Factor analysis: A means for theory and instrument development in support of construct validity. *International Journal of Medical Education*, 11, 245–247. <https://doi.org/10.5116/ijme.5f96.0f4a>
- Tideman, E., & Gustafsson, J.-E. (2004). Age-related differentiation of cognitive abilities in ages 3–7. *Personality and Individual Differences*, 36(8), 1965–1974. <https://doi.org/10.1016/j.paid.2003.09.004>
- Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology*, 45(4), 1097–1118. <https://doi.org/10.1037/a0015864>
- van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, 113(4), 842–861. <https://doi.org/10.1037/0033-295x.113.4.842>
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41(3), 321–327. <https://doi.org/10.1007/BF02293557>
- Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, 44(3), 219–246. <https://doi.org/10.1177/0095798418771807>
- Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods*, 1(4), 354–365. <https://doi.org/10.1037/1082-989X.1.4.354>
- Zimprich, D., & Martin, M. (2010). Differentiation-dedifferentiation as a guiding principle for the analysis of lifespan development. In A. Kruse (Ed.), *Life in old age: Personal and shared responsibility in society, culture and politics* (pp. 29–38). AKA.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.