American Psychological Association

Journal of Comparative Psychology

0735-7036/19/$12.00

2019, Vol. 133, No. 1, 92–105
http://dx.doi.org/10.1037/com0000136

The Structure of Ape (*Hominoidea*) Intelligence

Allison B. Kaufman
University of Connecticut, Avery Point

Matthew R. Reynolds
University of Kansas

Alan S. Kaufman
Yale University School of Medicine

We used contemporary psychometric theory of intelligence and confirmatory factor analysis to reanalyze data obtained on samples of nonhuman primates administered the Primate Cognition Test Battery. Our main goals were to interpret stability of the Primate Cognition Test Battery tasks and factors over time and to determine whether the cognitive factors that emerge from confirmatory factor analysis for apes can be interpreted from the perspective of a major theory of human intelligence, namely, the Cattell–Horn–Carroll model. We also analyzed data for 2½-year-old children on Wechsler's preschool test to afford a comparison between ape and child cognitive factors. Results indicated that multiple cognitive abilities provide the best factor solutions for both apes and children, and that the ape factors can be meaningfully interpreted from Cattell–Horn–Carroll theory.

Keywords: Primate Cognition Test Battery, Cattell–Horn–Carroll theory, Wechsler Preschool and Primary Scale of Intelligence-IV, nonhuman primate, psychometric

In humans, *g* or a general intelligence factor (Spearman, 1904, 1927) has enjoyed a lengthy, and often controversial, research history (Jensen, 1998). Although currently most human intelligence researchers and clinicians favor multiple abilities theories of cognitive development, *g* theory has yet to fall completely out of favor (Canivez, 2013; Canivez, Watkins, & Dombrowski, 2017).

The measurement of human intelligence began with Sir Francis Galton's (1869, 1883) perception/sensation approach and Alfred Binet's ingenious cognitive methodology (Binet & Simon, 1904), followed closely by Spearman's (1904, 1927) general intelligence or *g* theory. Soon after, David Wechsler (1939) moved beyond *g* when measuring human intelligence, introducing the verbal–performance (nonverbal) dichotomy as well as profiles of specific abilities; Wechsler's tests and clinical approach to children's and adults' assessment continue worldwide to the present day (Kaufman, Raiford, & Coalson, 2016). The field of IQ testing has witnessed the development of theory-based tests (Kaufman & Kaufman, 1983; Woodcock, McGrew, & Mather, 2001) that have been constructed with sophisticated psychometrics such as Rasch latent-trait models, confirmatory factor analysis (CFA), and structural equation modeling, and it has weathered numerous genetic controversies (Herrnstein & Murray, 1994; Rushton, 1997). Contemporary IQ testing has blossomed in the wake of dynamic, well-researched multiability theories that combine the notion of a general overall intelligence with smaller, specific intellectual abilities (Carroll, 1993; Horn & Cattell, 1966; Schneider & McGrew, 2012, in press).

This brief history of human intelligence stands in direct contrast to the history of nonhuman animal intelligence. Whereas human intelligence researchers started with the construct of *g* and worked outward toward smaller, multiple intelligences, nonhuman animal researchers have traditionally taken the opposite route—intelligence was seen from an evolutionary perspective and viewed as modular, dealing with specific skills and abilities (Balda, Kamil, & Bednekoff, 1996; Bond, Kamil, & Balda, 2003; MacLean et al., 2012). Pioneers in human intelligence testing, such as Binet, developed tasks that were cognitive (rather than social) in nature, that were intended to measure aspects of *g* or global capacity, and that could be administered in a laboratory or clinical setting. By contrast, the evolutionary approach taken by researchers of nonhuman animal intelligence viewed the mind as comprising domain-specific processes that were targeted to solve problems within their natural habitat—for example, related to foraging patterns and group living. This emphasis on the evolution of nonhuman animal intelligence led to insightful studies of differences among species

---

Allison B. Kaufman, Departments of Ecology and Evolutionary Biology and Psychology, University of Connecticut, Avery Point; Matthew R. Reynolds, Department of Educational Psychology, University of Kansas; Alan S. Kaufman, Child Study Center, Yale University School of Medicine.

We are extremely grateful to William Hopkins and Jamie Russell of the Hopkins Lab at Georgia State University for allowing us to use their data for this project. Many additional thanks to Jamie Russell for patiently answering our questions. We are also indebted to Josep Call and Esther Herrmann, original developers of the Primate Cognition Test Battery, for allowing us to use their instrument, and their data, and for their patience in answering our questions to clarify Primate Cognition Test Battery tasks. Finally, Allison B. Kaufman would like to thank James C. Kaufman for answering even more questions when no one else was available.

Correspondence concerning this article should be addressed to Allison B. Kaufman, Departments of Ecology and Evolutionary Biology and Psychology, University of Connecticut, Avery Point, 1084 Shennecossett Road, Groton, CT 06340. E-mail: allison.kaufman@uconn.edu

in solving visual-spatial problems (Balda et al., 1996), and applying reasoning to social situations (Bond et al., 2003; Isden, Panayi, Dingle, & Madden, 2013; MacLean et al., 2012). These studies, and others like them, have provided strong evidence that some species have particular expertise in specific domains, and that this kind of expertise is independent of their other cognitive abilities. Thus, although early tests and theories of human intelligence stressed a single global ability and were largely devoid of social context, early theories and tests of nonhuman animal intelligence emphasized separate domains and were rooted in social living and social problem solving.

Very recently, there has been a flurry of interest among nonhuman animal researchers on the *g* factor (see Burkart, Schubiger, & Van Schaik, 2016, for an excellent review of the literature on the existence and evolution of *g* in nonhuman animals). As the construct of *g* became more commonplace in the human intelligence research community, the first discussions of a nonhuman *g* began to occur in reference to mice and rats (Plomin, 2001). Of numerous studies that tested the intelligence of mice and rats in the past 25 years, most showed evidence of *g* (Anderson, 1993; Galsworthy et al., 2005; Kolata, Light, & Matzel, 2008; Kolata et al., 2005; Locurto & Scanlon, 1998; Matzel et al., 2003). Likewise, the bird literature has validated the *g* factor (Shaw, Boogert, Clayton, & Burns, 2015), although the main goal of many bird studies has been to correlate specific areas of intelligence with specific evolutionary behaviors—for example, problem-solving ability and mating success (Isden et al., 2013) or cognitive performance and song learning (Boogert, Anderson, Peters, Searcy, & Nowicki, 2011). The dog literature has also identified a *g* factor (Arden & Adams, 2016), as well as an array of domain-specific abilities (Maclean, Herrmann, Suchindran, & Hare, 2017).

In the ape literature, most researchers have tried to identify *g*, sometimes validating the general factor (Deaner, van Schaik, & Johnson, 2006; Reader, Hager, & Laland, 2011; Vonk & Povinelli, 2011), sometimes not (Herrmann & Call, 2012; Herrmann, Hernández-Lloreda, Call, Hare, & Tomasello, 2010). A more prevalent finding has been the emergence of multiple abilities that span the cognitive and social spheres (Herrmann & Call, 2012; Hopkins, Russell, & Schaeffer, 2014), consistent with evolutionary hypotheses that guided the early theoretical foundations of nonhuman animal intelligence.

In general, when apes and children are compared, it seems that apes are similar to 2- and 2½-year-old children in tests of spatial relations, causality, and quantities (i.e., physical tasks), but children excel in social tasks (intentionality, communication, and social learning). This ability to use social and cultural behaviors to learn and develop higher cognitive skills has been termed the *cultural intelligence hypothesis* (Herrmann, Call, Hernàndez-Lloreda, Hare, & Tomasello, 2007). Cultural intelligence leads to additional improvement of abilities in the physical world for humans that go beyond those of apes (Tomasello & Herrmann, 2010).

The cultural intelligence hypothesis was originally developed as a result of work with the Primate Cognition Test Battery (PCTB; Herrmann et al., 2007), which spans both cognitive and social domains. Variations and adaptations of the PCTB have since been used several times to measure intellectual abilities in apes (Herrmann & Call, 2012; Hopkins et al., 2014; Woodley of Menie, Fernandes, & Hopkins, 2015). The test is complex enough and includes enough of a variety of tasks as to allow more elaborate statistical analysis, and parallels the subtest format used in human intelligence tests (Kaufman, 2009). In addition, the potential for comparative analysis has been extended even more with the recent publication of the Dog Cognition Test Battery (Maclean et al., 2017).

Data are available on human children aged 2½ years, an age young enough (i.e., not yet saturated with cultural intelligence) to permit direct comparison with apes of the abilities measured by the PCTB (Herrmann et al., 2010). Of particular interest is to compare the factor structures of cognitive abilities for young children and apes and determine whether intelligence tests intended for apes yield factor structures that are interpretable via the same theories applied to intelligence tests for preschool children.

Cattell–Horn–Carroll Theory

The most popular multiple ability model that has been applied to the development and interpretation of intelligence tests for humans is the Cattell–Horn–Carroll (CHC) theory, which began with the work of Raymond Cattell (1941, 1963), who initially divided *g* into two types: Fluid Intelligence (Gf), the ability to solve novel, on-the-spot problems by using reasoning, with little dependence on previous learning; and Crystallized Intelligence (Gc), a knowledge-based ability that is dependent on education and acculturation and reflects "the depth and breadth and of knowledge and skills valued by one's culture" (Flanagan, Ortiz, & Alfonso, 2013, p. 17). Cattell's student, John Horn (1968), described four additional abilities—*Gsm* (Short-Term Memory), *Glr* (Long-Term Memory and Retrieval), *Gs* (Processing Speed), and *Gv* (Visual Processing—generating, perceiving, analyzing, manipulating, and transforming visual patterns and stimuli), with additional abilities added later. Gf–Gc theory was ultimately merged with three-stratum theory developed by John Carroll (1993), so that under an overall umbrella of *g* were ~10 broad intelligence abilities, further subdivided into ~70 narrower abilities; recent estimates place the number of broad abilities at 16, accompanied by 80–90 specific narrow abilities (Schneider & McGrew, 2012).

Each broad ability is further subdivided into specific narrow abilities. For example, Fluid Reasoning, or *Gf*, measures the narrow abilities of Induction (I), which involves discovering underlying principles or rules; General Sequential Reasoning (RG) is reasoning logically using known premises and principles; and Quantitative Reasoning (RQ) requires solving problems that involve numbers (Flanagan et al., 2013, p. 389). The CHC model has three levels (*g*, broad abilities, and narrow abilities), consistent with John Carroll's hierarchical model of intelligence; notably, however, John Horn was a firm *disbeliever* in *g* as a meaningful construct, and most proponents of CHC theory emphasize broad and narrow abilities while ignoring *g*.

The majority of modern IQ tests are written based on, and interpreted within, the framework of CHC theory (Kaufman, 2009). As a result, despite numerous IQ tests being available, results are comparable from one test to another (Reynolds, Keith, Flanagan, & Alfonso, 2013). In addition, CHC theory rests on an accumulation of a wealth of literature that attests to the construct validity of the broad abilities and narrow abilities, especially the ones that are measured by the major IQ tests for children, adolescents, and adults (Kaufman et al., 2016). Nonhuman data have not, to our knowledge, been examined via this lens, although there are parallels. For example, Herrmann and Call (2012) noted that their first factor was composed of learning tasks and the third factor was defined by inferential tasks. From the perspective of CHC theory, these factors reflect *Glr* and *Gf*, respectively (Kaufman, Reynolds, & Kaufman, 2013). The main goals of our study were (a) to apply the psychometric procedures used routinely in the study of human intelligence to data obtained on chimpanzees in recent large-scale studies (Hopkins et al., 2014), and (b) to determine whether factors obtained for chimpanzees via CFA are invariant over time and can be meaningfully interpreted from the vantage point of CHC theory (Schneider & McGrew, 2012, in press).

Goals and Setting for This Article

The field of nonhuman animal intelligence has increased rapidly in its sophistication regarding the interpretation of *g* factors (Woodley of Menie et al., 2015) and the application of CFA to data from nonhuman animals (Arden & Adams, 2016; Herrmann et al., 2010). Nonetheless, more integration of the fields of human and nonhuman animal intelligence regarding statistical methodology, standard psychometric practices, hierarchical models of intelligence that include *g* and specific broad abilities, and the desirability of applying CHC theory to the ability factors identified for apes may provide more insight into ape intelligence as measured by the PCTB.

Recent large-scale investigations of apes (Herrmann et al., 2010; Hopkins et al., 2014; Woodley of Menie et al., 2015) have demonstrated remarkable sophistication in the subtle modifications and enhancements of the PCTB. Here we wanted to examine some of the psychometric properties of the PCTB, such as the internal consistency and score stability, and to explore a few hypotheses about the latent constructs that underlie the PCTB using CFA. Further, using CFA, we wanted to test: (a) whether the same constructs are measured across time (i.e., equality of factor loadings across time), and (b) the stability of the constructs over time (i.e., magnitude of factor correlations from Time 1 to Time 2). We also wanted to attempt to interpret the constructs from a CHC lens and try to compare those with intelligence constructs measured in human toddlers. The Hopkins lab kindly provided us with their data (Hopkins et al., 2014; Lacreuse, Russell, Hopkins, & Herndon, 2014; Russell, Lyn, Schaeffer, & Hopkins, 2011), as did Esther Herrmann and Josep Call (Herrmann & Call, 2012; Herrmann et al., 2010), permitting us to expand upon their excellent work with the PCTB.

We analyzed data that Hopkins and colleagues collected from chimpanzees who were tested at two separate times, data on chimpanzees and 2½-year-old children tested on the PCTB (Herrmann et al., 2010), and data on apes tested on the PCTB (Herrmann & Call, 2012). We asked the following research questions:

(1) How reliable (internally consistent) are the PCTB composite scores, as modified by Hopkins and colleagues, and how stable are the scores earned by chimpanzees on the 13 separate tasks that compose the PCTB across an interval of 2–3 years?

(2) Does the factor structure of the PCTB, derived from a variety of CFA models, accord well with the principal components analysis (PCA) factor structure reported by Hopkins and colleagues for the 86 chimpanzees tested twice? Do CFA results accord well with the CFA factor structure reported by Herrmann et al. (2010) for 106 chimpanzees and the PCA structure reported by Herrmann and Call (2012) for 32 apes?

(3) Is the factor structure of the PCTB invariant (stable) from Time 1 to Time 2 (Hopkins et al., 2014), based on structural equation modeling/CFA, for the 86 chimpanzees tested twice, and how stable are the factors? Is the factor structure invariant for the chimpanzees and 2½-year-old children (Herrmann et al., 2010)?

(4) Can the factors identified by CFA for apes be interpreted within the context of CHC theory, which has previously been developed and validated for humans?

(5) How similar are the factors identified for chimpanzees in the present study to the factors identified (also in the present study) for 200 children, aged 2½ years (2:6–2:11), tested on the Wechsler Preschool and Primary Scale of Intelligence-IV (WPPSI-IV) during the standardization of that children's IQ test?

Method

Sample

The primary set of data used were collected on apes by William Hopkins and colleagues (Hopkins et al., 2014; Lacreuse et al., 2014; Russell et al., 2011; Woodley of Menie et al., 2015) at a variety of facilities—the Jacksonville Zoo, Milwaukee Zoo, The Great Ape Trust of Iowa, Language Research Center (Atlanta, Georgia), and the Yerkes National Primate Research Center (Atlanta, Georgia). Eighty-six chimpanzees were tested with a modified version of the PCTB (Herrmann et al., 2007) and then retested ~2 years later. All were from Yerkes Center with 37 mother reared, 39 nursery reared, and 10 wild; of the group retested, 62 were females. We also analyzed PCTB data collected by Herrmann et al. (2010) on chimpanzees ($N = 106$) and 2½-year-old children ($N = 105$), by Herrmann and Call (2012) on 32 apes (14 chimpanzees, seven gorillas, seven orangutans, and four bonobos), and data collected on 200 children aged 2½ on the WPPSI-IV (Wechsler, 2012) during the standardization of that human IQ test.

Analysis Plan

Reliability and stability of the 13 PCTB tasks and composite. Correlations were computed between the chimpanzees' scores on the individual tasks at Time 1 and Time 2. These correlations represented stability coefficients. In addition, mean scores and Cohen's *d* effect sizes, accounting for the correlations across time (Morris & DeShon, 2002), of score differences across the two time points were calculated. An overall composite of the tests was also created by standardizing the scores on each individual task and taking the mean of those *z*-scores. Cronbach's α estimates for these composites were calculated at each time point, in addition to the correlation, mean difference, and effect size.

Determining CFA factor structure of the PCTB. CFA models were tested with Time 1 data to investigate plausible constructs that underlie the PCTB. Models were derived from previous research with these data, previous research using PCTB tests with children and apes, and from our knowledge of CHC theory with humans.

PCTB, PCA, and hierarchical scoring structure. We used CFA to test the plausibility of a structure based on the number of components and loadings that were statistically significant in the PCA (Model 1 in our study) reported in Hopkins et al. (2014). In their study, Hopkins et al. (2014; see SF1) also created different composites by averaging different configurations of the PCTB tests and reported heritability estimates for those composites in the supplemental materials. The structure of the PCTB scores was represented as a hierarchy, with an overall score at the top (composite of all of the tests), and physical and social scores below that level. Scores for *causality*, *space*, and *number discrimination* were calculated (physical domain), as were scores for *communication* and *theory of mind* (social domain). Here we imposed a CFA structure similar to the authors' hierarchical scoring structure as a way to assess the validity of those constructs. We tested three models based on their scoring structure: (a) *five-factor model* (Model 2), (b) *five-factor model, with higher order Physical factor* (Model 3), and (c) *two-factor model* (Model 4), validated in previous CFA with PCTB data (Herrmann et al., 2010). In addition, we tested a *three-factor model* (Model 5), derived from a three-factor model that Herrmann et al. (2010) found in a CFA with data from 2-year-old *children* who were administered the PCTB battery.

CHC structure on PCTB. The last set of models included those that were most consistent with our knowledge of CHC theory, and ones that formally included a *g* factor. The first model was a correlated first-order factor model (Model 7). Four correlated first-order factors were included: We described Social as most closely associated with Gc, Space as associated with Gv, and the other two factors as most similar to Gf (but one that required the mechanical use of tools). We also specified a slight variation of this model with a correlation between the attention state and production residuals (Model 8).

We specified a higher order model (Model 9) that included a second-order *g*-factor to explain, in part, the covariances among the first-order factors that were established in Models 7 and 8. This model is particularly useful because the second-order factor loadings may be interpreted in a way to understand the *g* factor (Keith & Reynolds, 2012). Notably, Gf factors are typically most closely associated with a *g* factor in humans (Carroll, 1993).

We specified a bifactor model (Reise, 2012). In the bifactor model (Model 10), a general factor is included as a first-order factor and is associated directly with every test. In addition, broad ability factors are also included (e.g., Gc). These broad ability factors represent factors *where g* has first been removed from all of the tests. All of these factors are uncorrelated. We specified a model with a single *g* factor (Model 11) only to test whether all test covariances were adequately accounted for by a single *g* factor (Spearman, 1904, 1927).

CFA of other samples of apes. We tried to cross-validate our best CFA models obtained on Hopkins' 86 chimpanzees with the 106 chimpanzees tested by Herrmann et al. (2010), who also conducted CFA. We also conducted CFA on the ape data ($N = 32$) on which Herrmann and Call (2012) extracted three components using PCA.

Stability of factor structure for 86 chimpanzees tested twice. The best-fitting and most plausible CFA models from Time 1 were applied to Time 2 data. The purpose was to evaluate whether the models were configurally the same across time—whether the same factors were found on each separate administration of the PCTB.

The CFA model that generalized across the time points was used in a longitudinal multiple group model; here we tested for factorial invariance, specifically for factor loading invariance. To do so, corresponding factor loadings were fixed (i.e., predetermined or "set") to be equal to each other across time. Model fit was then compared with a model in which the loadings were not fixed equal—a test of whether the constructs were essentially the same across time.

Factor structure of WPPSI-IV at age 2½. CFA was conducted on data from the 200 children (100 boys, 100 girls) aged 2½ years included in the WPPSI-IV standardization sample. The correlation matrix with standard deviations for the seven subtests for ages 2½ (Wechsler, 2012; Table A.1, p. 158)[1] was converted to a covariance matrix and subjected to a correlated three-factor CFA.

Model evaluation. For all CFAs, the root mean square error of approximation (RMSEA) and the comparative fit index (CFI) were used to evaluate fit of standalone models. RMSEA values close to or less than .05 and CFI values close to or greater than .95 are often used as indicators of "good" global fit (cf., Hu & Bentler, 1999; Keith, 2015; Schermelleh-Engel, Moosbrugger, & Müller, 2003). Model chi-square was also interpreted and whether the fit was statistically significantly different from zero. The likelihood ratio test ($\Delta\chi^2$) and Akaike's information criterion were used to evaluate competing and alternative plausible models; $\Delta\chi^2$ was used for nested model comparisons. In addition to global evaluation of the models using fit statistics and indexes, we examined the plausibility of the estimates and searched for any evidence of local misfit. We also evaluated the sufficiency of the factor loadings to evaluate the viability of the factors.

Results

Reliability and Stability of the PCTB

Table 1 shows Time 1 and Time 2 mean PCTB scores for the 86 chimpanzees (mean interval = 2.43 years). As shown, nine means increased at Time 2, three decreased, and one stayed the same. Cohen's *d* effect sizes for three mean increases were ≥.30: relative numbers, object permanence, and rotation.

Time 1 and Time 2 stability correlations are shown in the column furthest to the right in Table 1. The correlations are less than those typically observed in humans. For the 13 PCTB tests, stability coefficients ranged from −.09 (for causality—visual) to .48 (transpose; median coefficient = .24). By contrast, for 2½-

---

[1] Although CFA of the WPPSI-IV was published in the Technical Manual for ages 2 years, 6 months to 3 years, 11 months and 4 years to 7 years, 7 months (Wechsler, 2012, Figures 5.1 & 5.2), analyses were not conducted for separate age-groups within that range. For this study, we performed CFA on the 200 children ages 2 years, 6 months to 2 years, 11 months to provide the best comparison of factor structure to the CFA structure obtained for the sample of 86 chimpanzees.

Table 1
Time 1 and Time 2 Primate Cognition Test Battery (PCTB) Means, Standard Deviations, Mean Differences, and Stability Correlations

	Time 1		Time 2		Time 2–Time 1		Time 1 Time 2
PCTB measure	*M*	*SD*	*M*	*SD*	*M*	Cohen's *d*s	*r*s
Tool use	.37	.49	.43	.50	.06	0.11	.40
Relatives	.65	.15	.71	.15	.06	0.30	.08
Causality noise	.51	.13	.55	.16	.03	0.20	.06
Causality visual	.67	.17	.64	.17	−.03	−0.12	−.09
Tool properties	.66	.21	.66	.18	.00	0.00	−.05
Transpose	.59	.24	.65	.24	.06	0.25	.48
Spatial memory	.59	.28	.62	.26	.03	0.08	.13
Rotation	.44	.18	.53	.21	.09	0.36	.17
Object permanence	.61	.18	.70	.17	.09	0.47	.40
Comprehension	.46	.30	.44	.29	−.03	−0.07	.45
Attention state	.40	.27	.44	.31	.04	0.11	.24
Gaze	.60	.37	.48	.37	−.12	−0.27	.31
Production	.38	.39	.44	.40	.06	0.14	.37

Note. $N = 86$, except causality noise ($N = 83$). Average time in between testing = 2.43 years, $SD = 1.34$; range = 5 years (min = 1; max = 6). Negative values for means and Cohen's *d*s indicate lower performance at Time 2.

year-old children on the WPPSI-IV, coefficients ranged from .72 to .84 (median = .78; Wechsler, 2012; Table 4.5), although the interval was 3 weeks, not 2 years. Tests for humans include many items (or behavioral samples), unlike the measures used here.

The Time 1 and Time 2 stability correlation for the composite of PCTB tasks was .50, and on average the scores improved ($d = .18$). Coefficient alpha estimates of .54 for Time 1 and .56 for Time 2 were observed for the composites. On the WPPSI-IV at age 2½, Full Scale IQ had a stability coefficient of .90 and a reported measure of internal consistency of .96 (Wechsler, 2012; Tables 4.1 & 4.5).

CFA Structure of the PCTB for Hopkins's Chimpanzees

Table 2 summarizes the CFA results for the 11 models (see Method section) for the 86 chimpanzees tested twice by Hopkins et al. (2014). Four models (Models 6, 8, 9, and 10) provided good fits for Time 1 data (also see Figures 1 and 2). Model 6 is important because this model was derived from data on *children* (Herrmann et al., 2010) that is now shown to be applicable to chimpanzees; also, this was the only good-fitting model *not* derived from human CHC theory.[2] Bifactor Model 10 yielded the best fit indexes. Both Models 8 and 9 validate the four separate CHC-based abilities for chimpanzees—the Gf factors were absorbed into the *g* factor in the bifactor model. The higher order Model 9 provides cross-validation of the *g* factor in chimpanzees, as does the bifactor model. Any of the four models could be chosen as the best. We have illustrated with Model 9 (see Figure 1), which has excellent fit data and includes four meaningful CHC factors, and bifactor Model 10 (see Figure 2). By contrast, the models derived from Hopkins et al.'s (2014) PCA and hierarchical scoring system (Models 1–4) produced unsatisfactory fits.

Model 11, which hypothesized *only* a *g* factor, provided a poor fit, much worse than most other models. However, the existence of a solid *g* construct for chimpanzees is validated by the results of the hierarchical CHC models (Models 9 and 10), both of which showed a robust *g* factor.

CFA of the PCTB for Herrmann et al.'s Chimpanzees and Herrmann and Call's Apes

When we conducted CFA on Herrmann et al.'s (2010) 106 chimpanzees, we ran into the same problem they confronted with their CFAs—most models did not converge, including a *g*-only model and our CHC-based multiability models (when models do not converge, one cannot compute fit statistics). We were able to get convergence on the two-factor model they reported and also on a bifactor model. The bifactor model yielded a *g* factor and a Gv factor, but almost all of the loadings were not statistically significant. The RMSEA was excellent, but the CFI was barely adequate. Herrmann et al.'s best CFA model for chimpanzees also showed low loadings on their two factors (Physical/Social and Spatial—five of seven loadings were ≤.17). Overall, our models for the 86 Hopkins chimpanzees did *not* generalize to Herrmann's 106 chimpanzees.

The CFA based on the PCA of seven Physical tasks analyzed by Herrmann and Call (2012) for 32 apes yielded three correlated factors (Gf, Glr, and Gv in CHC parlance). The PCTB battery included three discriminative learning tasks (Shape, Size, and Color), which have generally been excluded from other versions of the PCTB (Herrmann & Call, 2012; Hopkins et al., 2014). These three tasks joined with Space to form Factor 2 (labeled Long-Term Retrieval or Glr). The fit statistics were marginal; nonetheless, chi-square was not significantly different from zero, which suggests an adequate fit to the data, and the three correlated factors identified by the CFA resemble closely the three uncorrelated factors reported by Herrmann and Call (2012) based on PCA (the factor correlations ranged from .12 to .36 in the CFA). The hierarchical model had the same fit statistics as the three-factor model because they are statistically indistinguishable, and both fit better than the *g*-only model.

Invariance of the CFA Factor Structure for Chimpanzees From Time 1 to Time 2

Three of the best Time 1 models (Models 6, 9, and 10) were tested with data from Time 2. The Physical and Tool factors (reasoning ability), whether or not these were combined into a single factor, were not viable in these models. The Space factor was easy to identify for the most part and Social/Communication worked in most models.

A bifactor model with Time 2 data was successfully estimated. Although the models cannot be compared directly, the fit of this bifactor model for Time 2 data, based on standalone fit indexes, was worse, $\chi^2(59) = 80.90$, $p = .03$, RMSEA = .07, CFI = .80, than with Time 1 data. The most salient finding was that the *g*

---

[2] Note that a better fit (Model 6) was obtained by freeing a residual correlation between attention state and production (freeing means that a correlation was allowed, it was "freed" from being set to zero). This finding suggests that another factor is needed to account for common variance in these tests beyond that accounted for by the Social/Communication factor.

Table 2
Fit of Confirmatory Factor Analysis Models With the Primate Cognition Test Battery Time 1 Data

Model	χ^2	*df*	*p*	$\Delta\chi^2$	Δdf	*p*	CFI	RMSEA	AIC
1. Hopkins et al.'s[a] PCA model	92.32	63	.01				0.59	0.07	174.3
2. Hopkins et al.'s five-factor	70.72	57	.11				0.81	0.05	164.7
3. Hopkins et al.'s five-factor H-O	75.08	61	.11				0.81	0.05	161.1
4. Hopkins et al.'s two-factor	80.09	64	.08				0.77	0.05	160.1
5. Herrmann's[b] three-factor w/ children	72.81	63	.18				0.86	0.04	154.8
6. Herrmann's three-factor w/ residual	68.72	62	.26	4.09	1	.04	0.91	0.04	152.7
7. CHC	68.83	60	.20				0.88	0.04	156.8
8. CHC w/residual	62.73	59	.35	6.10	1	<.05	0.95	0.03	152.7
9. CHC H-O/residual	66.46	61	.29	3.73	2	.15	0.92	0.03	152.5
10. Bifactor	56.51	59	.57				1.00	<.01	146.5
11. Single *g* factor	92.17	65	.02	35.66	2	<.05	0.60	0.07	170.2

Note. Use previous model in Table for comparisons via $\Delta\chi^2$ when reported. CFI = comparative fit index; RMSEA = root mean square error of approximation; AIC = Akaike's information criterion; PCA = principal components analysis; H-O = higher order factor model; CHC = Cattell–Horn–Carroll.
[a] Hopkins et al. (2014). [b] Herrmann et al. (2010).

factor and Space (Gv) factor were the easiest to identify across time points.

Factor Invariance and Stability: Time 1 and Time 2

Because not all of the factors replicated across time, we only tested the invariance and stability of *g*, Space (Gv), and Social/Communication (Gc) factors across time using a bifactor model. These analyses involved estimating bifactor models for Time 1 data and for Time 2 data in the same analysis. Correlated residuals among corresponding tests were allowed across time. Because the tests have exactly the same content, they likely share variance beyond variance they share with other tasks within each time point. The fit of model was adequate according to the RMSEA, but not to the CFI (see Table 3). The factor loadings were invariant across time, suggesting the constructs were the same across time. In the *all invariant model* with all corresponding factor loadings equal across time, the factor correlations were as follows: .75 for *g*, .68 for Gv, and .64 for Gc. These correlations quantify the stability of the factors from Time 1 to Time 2 and are good, although stability coefficients in humans are typically .80 or greater for specific abilities like Gv and are usually .90 or greater for *g* factors (Kaufman et al., 2016).

Factor Invariance Chimpanzees Versus Children

We were unsuccessful in finding models that converged when trying to estimate the factor structures of chimpanzees and children using the covariance matrices from the Herrmann et al. (2010) study. Therefore, rather than test the factor structure, we tested whether the constructs were similar across groups, without defining what those constructs are, by testing the equivalence of the covariance matrices across chimpanzees and children. The two were *not* equivalent, although the misfit was mostly due to differences in variance across the social tasks.

Analysis of Ape Factors From the Vantage Point of CHC Theory

To evaluate whether or not the PCTB subtests and factors can be interpreted from the vantage point of CHC theory, we relied on the sources shown in the *Note* to Table 4. Based on our CHC classification of PCTB tasks, we would label the first-order factors in the four-factor CHC model as follows (see Figure 1): I—Fluid Reasoning-1 (Gf1), II—Fluid Reasoning-2 (Gf2), III—Visual-Spatial Processing (Gv), and IV—Crystallized Knowledge (Gc). The three Social/Communication tasks that loaded on the factor labeled Gc each has elements of Gc (see Table 4). In humans, Gc often involves language, but it also relates to communication ability and knowledge in general. Only Gaze Following was not associated with any of the factors. The fourth ape factor, composed of three social cognition tasks, may measure an aspect of crystallized knowledge (Gc); these tasks involve communication ability and dealing with practical (as opposed to emotional) information, a core aspect of the definition and conception of Gc (N. Mather, personal communication, August 8, 2016; J. Schneider, personal communication, August 9, 2016).

CFA Factor Structure for Children Aged 2½ on the WPPSI-IV

Figure 3 shows the factor structure for children aged 2½ on WPPSI-IV. The factors correspond to Gc, Gv, and Gsm/Gf, respectively. These factors are quite similar to the factors identified for the chimpanzees on the PCTB in this study and in other investigations of apes, with both measuring similar Gv components and an "ape" Gc, which appears to target communicative abilities as opposed to specifically verbal ones. Whereas Gf was not identified specifically for 2½-year-old children on the WPPSI-IV, the Working Memory factor bears a close kinship to Gf. Working memory—the ability to take in and hold onto information, keep it in immediate awareness, and use it within a few seconds—is an essential ingredient needed for solving novel problems; it shows strong ties to fluid reasoning (Burgess & Braver, 2010; Hornung, Brunner, Reuter, & Martin, 2011). Hence, we have used the label Gsm/Gf. Thus, the four ape factors (Gf1, Gf2, Gv, and Gc) correspond to broad ability factors identified for children as young as age 2½ on the WPPSI-IV and for adolescents and adults (Kaufman et al., 2016; Wechsler, 2012). Further, the factors identified for 2-year-old *children* administered the PCTB (Herrmann & Call, 2012) provide a good CFA model fit for

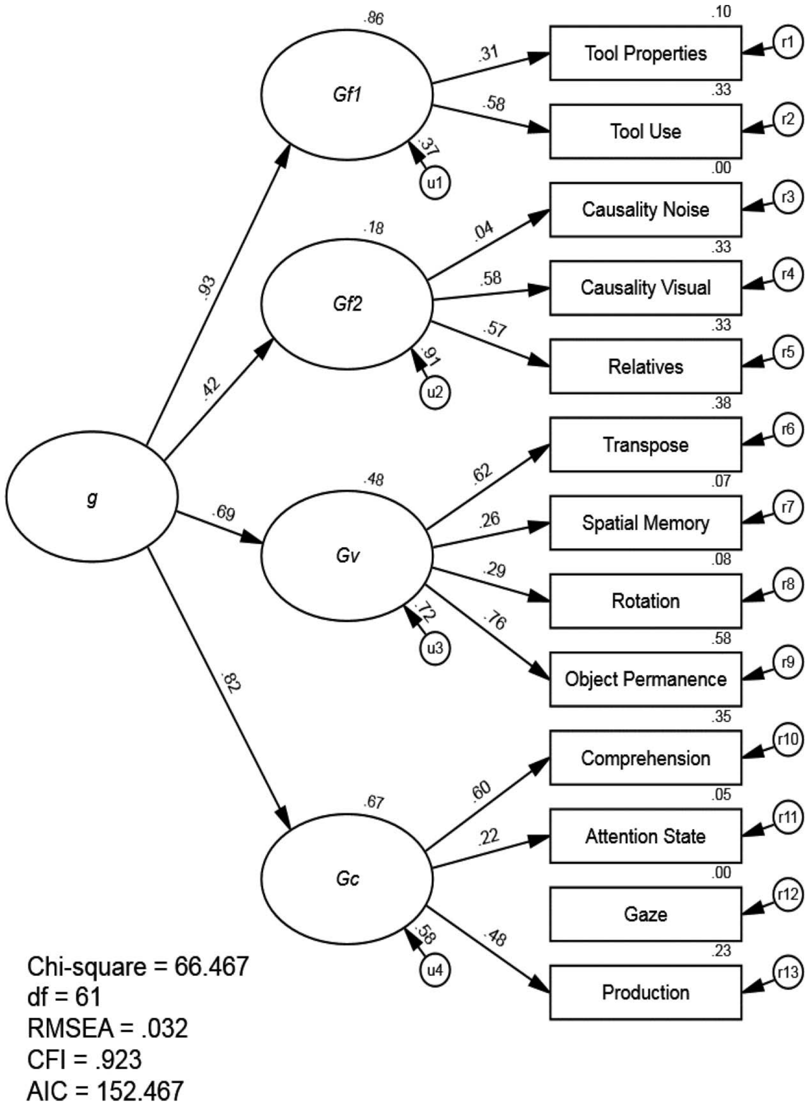

Figure 1. Cattell–Horn–Carroll confirmatory factor analysis Model 9 with Primate Cognition Test Battery Time 1 data for $N = 86$ (Hopkins et al., 2014). Rectangles represent the observed scores on each task. Ovals represent the latent variables. The standardized factor loadings are represented by the directed paths (single-head arrows). All estimates on the top right corner of each rectangle represents the squared multiple correlation or variance explained in the observed score by the common factor (i.e., communality). The ovals labeled r1–r13 represent the residual variance, or variance in the observed scores not explained by the common factor. A residual covariance between attention and production subtest was included, but not shown. The ovals labeled u1–u4 are the factor uniqueness. They represent the first-order factors with the common variance associated with the second-order *g* factor removed. The second-order *g* factor in this model influences all observed score indirectly through the first-order factors. The directed arrows from the unique variance to the first-order factors multiplied by the first-order factor loadings represent the unique effect of the first-order factor (i.e., effect of *g* is removed) on the observed score. To obtain the effects of *g* on observed scores, the second-order *g* factor loading is multiplied by the observed score loading on the respective factor. For example, the effect of *g* (i.e., g loading) on the Transpose subtest (.69 × .62) = .43. The unique effect of Gv on Transpose (.72 × .62) = .45. *g* = general intelligence factor; *Gf* = Fluid Intelligence; *Gv* = Visual Processing; *Gc* = Crystallized Intelligence; RMSEA = root mean square error of approximation; CFI = comparative fit index; AIC = Akaike's information criterion.

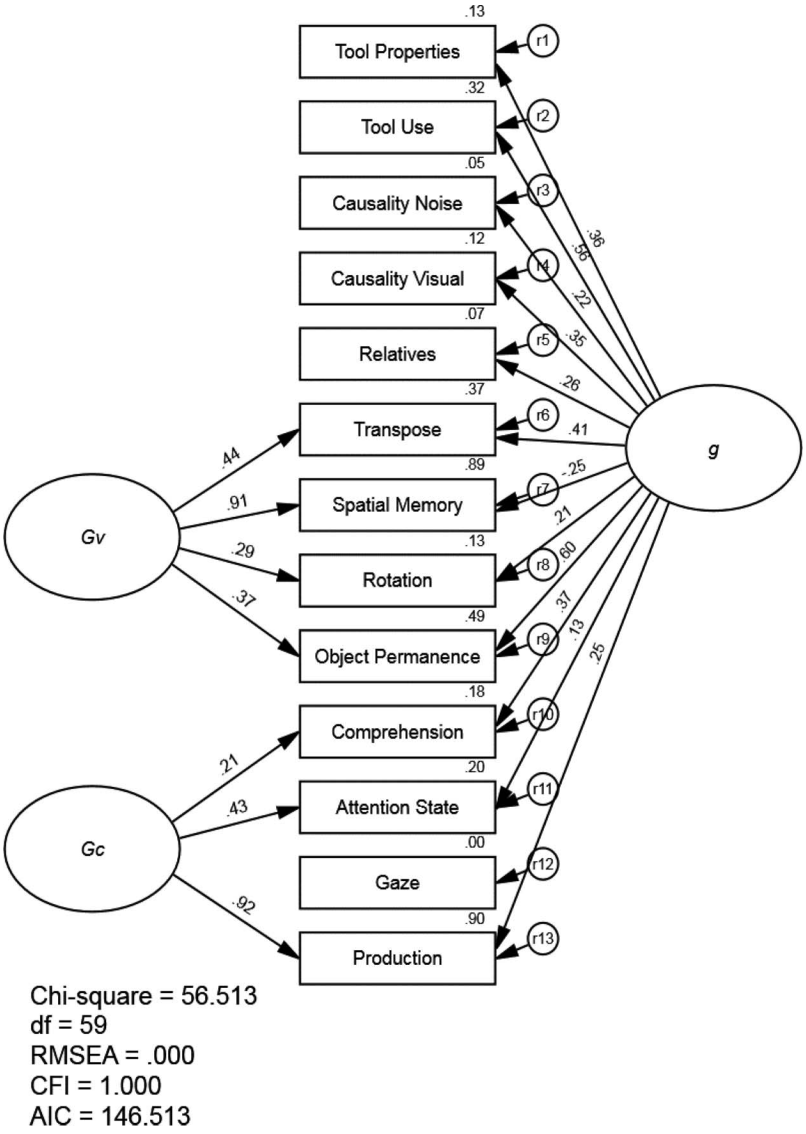

Figure 2. Bifactor Model 10 with Primate Cognition Test Battery Time 1 data for *N* = 86 (Hopkins et al., 2014). *g* = general intelligence factor; *Gv* = Visual Processing; *Gc* = Crystallized Intelligence; RMSEA = root mean square error of approximation; CFI = comparative fit index; AIC = Akaike's information criterion.

chimpanzees in the Hopkins et al. (2014) data set (see Model 6, Table 2).

Discussion

Interest has been building in the use of psychometric analysis on the data from nonhuman animals. In addition to the findings presented here, two reviews published in 2017 have summarized the information available, tracked the development and use of data, and identified areas where more work is needed. Matzel and Sauce (2017) have contributed to the literature by emphasizing the importance of understanding individual differences in performance on assessment, and pointing out how the information obtained from assessment can be applied to understand factors beyond intelligence such as personality, resilience, and mental health. Burkart et al. (2016) added to the literature by tracing the development of intelligence in social as well as nonsocial situations for both domain-general and domain-specific intelligences. Our goal was more specific. Our work builds on these two reviews by transitioning into empiricism and applying commonly used techniques to begin to understand if hypotheses reviewed in these and other articles can be supported with data. Our contributions to the literature are primarily the results of our psychometric and theory-based analyses that permit reinterpretation of and reanalysis of the constructs measured by the PCTB.

We were able to meaningfully interpret the results of the PCTB to find both a viable *g* factor and four separate ability factors in accordance with CHC theory, and which are also a good match for factors we identified for 2½-year-old children on the WPPSI–IV

Table 3
Time 1 and Time 2 Longitudinal Invariance Models

Model	χ^2	*df*	*p*	$\Delta\chi^2$	Δdf	*p*	CFI	RMSEA	AIC
1. Bifactor	360.0	281	.001				0.74	0.06	550.0
2. *g* loadings equal	376.0	294	.001	16.0	12	.19	0.724	0.057	541.0
3. Space loadings equal	363.3	285	.001	3.2	3	.36	0.737	0.057	547.3
4. Social/Communication loadings equal	365.6	284	.001	5.5	2	.06	0.726	0.058	551.6
5. All loadings equal	384.2	299	.001	24.2	17	.12	0.714	0.058	540.2

Note. Compare all models with Model 1. CFI = comparative fit index; RMSEA = root mean square error of approximation; AIC = Akaike's information criterion.

and the CFA factors that Herrmann and Call (2012) found for 2½-year-old *children* tested on the PCTB. The factors identified for chimpanzees represent Gv, Gf, and Gc (Figures 1 and 2) and also Glr (in the reanalysis of Herrmann and Call's [2012] data). Gc measures verbal abilities and accumulated knowledge and learning in children, and it measures the related concepts of general communication abilities and behavioral learning in apes. Though the *g* factor in chimpanzees is a viable construct (Figures 1 and 2), this factor alone provides a poor model fit to the ape data set (see Table 2). Similar to the psychometric structure of humans, the structure for nonhuman animals is likely multidimensional in nature, a finding observed by MacLean et al. (2017) in their analyses of the intelligence of dogs, using a test battery patterned on the PCTB. However, regardless of the positive findings we identified for chimpanzees and children, there was likewise a cautionary finding—the specific CHC factor structure identified for one sample of apes at a given point in time does not seem to fully generalize to the same sample when tested a second time (see Table 3) or to a different sample of chimpanzees from other laboratories.

Another main result of our study is the fact that the PCA conducted by Hopkins et al. (2014) with their chimpanzee data, as well as their assignment of tasks to domain-specific scales, was not supported by the results of our CFA (see Table 2). One advantage of CFA over PCA is that PCA is a data-reduction technique whereas factor analysis investigates the underlying latent structure. Further, CFA in particular, is useful in that specific hypotheses, such as assignment of tasks to domain-specific scales, can be tested empirically (see Herrmann et al., 2010). Exploratory factor analytic (EFA) techniques are useful for identifying the underlying structure of a test battery when that structure is unknown, which supports the variety of EFAs conducted by Hopkins and other nonhuman animal researchers. However, EFA methodologies are not as robust as confirmatory techniques, especially when the structure is known or is hypothesized to measure theory-based constructs. EFA relies on a purely psychometric approach to the data where the number of meaningful factors differs based on the criterion selected and the constructs measured by the factors are subjectively determined. The PCA method used by Hopkins and colleagues and by numerous other nonhuman animal researchers also has an additional disadvantage—the mathematical model places 1's in the diagonals, thereby making the untenable assumption that all tasks are perfectly reliable. By contrast, CFA (such as the excellent analyses conducted by Herrmann et al. [2010] on samples of chimpanzees and children) specifies the number of factors in advance and the constructs they are presumed to measure based on (a) previous validated research results, usually with (b) a viable theory of intelligence to guide the choice of proposed models. Also, CFA is accompanied by sets of criteria for judging goodness of model fit—criteria that can be applied validly to any data set.

The Glr factor identified for the 32 apes tested by Herrmann and Call (2012) measures the ability to learn new information during the testing session—in this case, learning to discriminate stimuli based on color, shape, and size. This factor has not emerged in other studies of apes because the three discriminative learning tasks in the original PCTB have been excluded from most subsequent versions. These PCTB learning tasks are particularly interesting because they are so similar to classical tests of long-term memory and retrieval in human IQ tests. Glr is not measured in the WPPSI-IV, but this factor has been identified for preschool children on the KABC-II (Kaufman Assessment Battery for Children-II; Kaufman & Kaufman, 2004) and on various versions of the Woodcock–Johnson (Schrank, McGrew, & Mather, 2014).

Other important results of our study are the application of structural equation modeling/CFA to determine the invariance of the overall factor structures at Time 1 and Time 2 for the 86 chimpanzees tested by Hopkins et al. (2014), and also for the chimpanzees and 2½-year-old children and chimpanzees tested on the PCTB by Herrmann et al. (2010). These results indicated a lack of invariance in the two factor structures in each data set. Nevertheless, the *g*, Gv, and Gc factors were invariant across time in the same sample, and the *g* factor was adequately stable ($r = .75$) as were the Gv ($r = .68$) and Gc ($r = .65$). The value of .75 for *g*, a latent variable, is substantially higher than the stability coefficient of .50 for the PCTB composite. In other words, scores on the global test score are somewhat, but not entirely stable (the separate tasks in particular—see Table 1) but the underlying constructs for *g*, Gv, and Gc are reasonably consistent over time. Much of the difference is likely explained by the elimination of measurement error in the construct correlations.

Implications for the Cultural Intelligence Hypothesis

The PCTB has been used to provide empirical support for the cultural intelligence hypothesis (Herrmann et al., 2007), and these results add to this growing body of data. Consistent with the evolutionary approach taken by nonhuman animal researchers, the PCTB and accompanying hypotheses include much more speculation on the development of social abilities in nonhuman animals. In humans, these skills are not typically the subject of

Table 4

Cattell–Horn–Carroll (CHC) Classifications of Primate Cognition Test Battery Tasks

- Fluid Reasoning (Gf) Tasks
  - Relative numbers:
    - Broad abilities—Gf, with some Quantitative Knowledge (Gq) and Gv as well
    - Main narrow abilities—General Sequential Reasoning (Gf), Induction (Gf), Visual Memory (Gv), Visualization (Gv), Mathematical Knowledge (Gq)
  - Causality noise:
    - Broad abilities—Gf
    - Main narrow abilities—General Sequential Reasoning (Gf), Quantitative Reasoning (Gf)
  - Causality visual:
    - Broad abilities—Gf
    - Main narrow abilities—Induction (Gf), General Sequential Reasoning (Gf)
  - Tool properties:
    - Broad abilities—Gf, with some Gv as well
    - Main narrow abilities—General Sequential Reasoning (Gf), Visualization (Gv)
  - Tool use:
    - Broad abilities—Gf, with some Gv and Psychomotor Abilities (Gp) as well
    - Main narrow abilities—General Sequential Reasoning (Gf), Visualization (Gv), Psychomotor Speed (Gp)
- Visual–Spatial Processing (Gv) Tasks
  - Spatial memory:
    - Broad ability—Gv
    - Main narrow abilities—Visual Memory (Gv), Spatial Scanning (Gv)
  - Object permanence:
    - Broad abilities—Gv, with some Gf as well
    - Main narrow abilities—Visual Memory (Gv), Perceptual Illusions (Gv), Piagetian Reasoning (Gf)
  - Rotations:
    - Broad ability—Gv
    - Main narrow abilities—Visualization, Speeded Rotation
  - Transposition:
    - Broad abilities—Gv, with some Gf as well
    - Main narrow abilities—Visualization (Gv), Visual Memory (Gv), Induction (Gf)
- Crystallized Knowledge (Gc) Tasks
  - Comprehension:
    - Broad abilities—Gc, with some Gv and Gf as well
    - Main narrow abilities—Knowledge of Behavioral Content (Gc), Communication Ability (Gc), Induction (Gf)
  - Production:
    - Broad abilities—Gc and Gf
    - Main narrow abilities—Communication Ability (Gc), Induction (Gf)
  - Attention state:
    - Broad abilities—Gc (perhaps)
    - Main narrow ability—Communication Ability (Gc)
- No clear CHC classification
  - Gaze following:
    - Broad abilities—Gv (perhaps)
    - Main narrow ability—Spatial Scanning (Gv)
- Long-Term Storage and Retrieval (Glr)—Tasks included only in Herrmann and Call (2012)
  - Shape:
    - Broad abilities—Glr with some Gv
    - Main narrow abilities—Associative Memory (Glr), Visualization (Gv)
  - Size:
    - Broad abilities—Glr with some Gv
    - Main narrow abilities—Associative Memory (Glr), Visualization (Gv)
  - Color:
    - Broad ability—Glr
    - Main narrow ability—Associative Memory (Glr)

Note. To develop this table, we relied on the following sources: (a) the classification and interpretation of CHC tasks on the Woodcock Johnson Tests of Cognitive Abilities–Fourth Edition (WJ IV; Schrank, Dekker, & Garruto, 2016; Schrank, McGrew, & Mather, 2014), (b) the cross-battery assessment system developed by Dawn Flanagan and her colleagues (Flanagan, Ortiz, & Alfonso, 2013), (c) recent authoritative chapters on CHC theory (Schneider & McGrew, 2012, in press), and (d) our own research (Niileksela, Reynolds, & Kaufman, 2013) and tests (Kaufman & Kaufman, 2004, 2018).

assessment on IQ tests, despite the fact that social learning in infants and young children plays a crucial role in human mental development (Piaget, 1936) as well as nonhuman animal mental development. Probably because tasks developed for humans originated to meet *practical* concerns such as school failure (Binet & Simon, 1904) or selection of draftees for the US Army (Yoakum & Yerkes, 1920), assessment of human intelligence rarely includes measurement of social capabilities that develop without formal

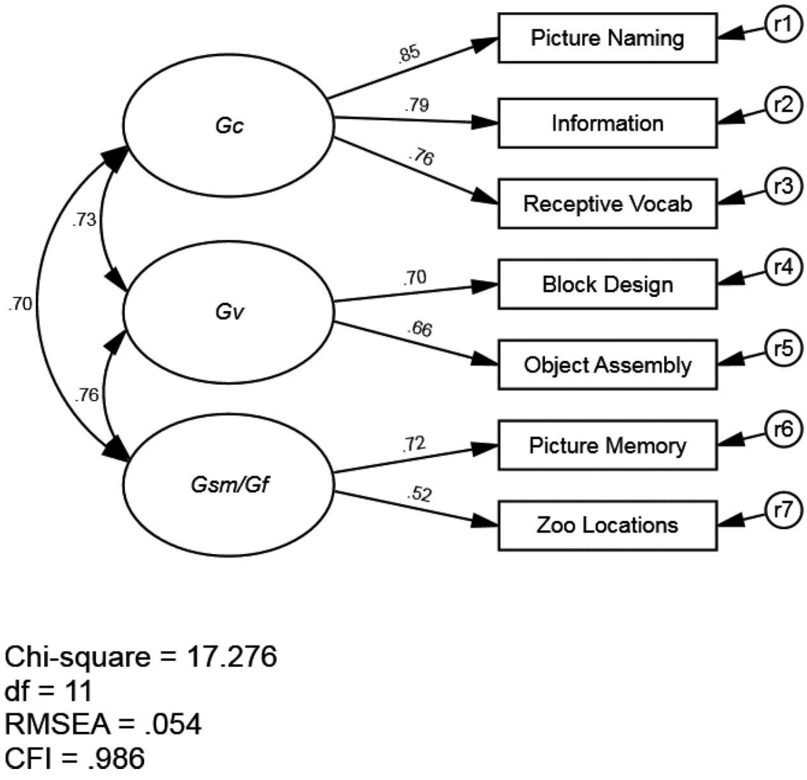

Figure 3. Confirmatory factor analysis of Wechsler Preschool and Primary Scale of Intelligence-IV data for 2½-year-olds (*N* = 200). *Gf* = Fluid Intelligence; *Gv* = Visual Processing; *Gc* = Crystallized Intelligence; *Gsm* = Short-Term Memory; RMSEA = root mean square error of approximation; CFI = comparative fit index.

education. Social intelligence is also not often measured because it is difficult to develop standardized social tasks for humans that can be administered within a laboratory or clinical setting and scored objectively. In addition, IQ test construction for humans has emphasized psychometric considerations (how well does a task measure *g*?) rather than theoretical factors (how does an infant's or child's intelligence develop?) This psychometric emphasis is evident in the 16 or so CHC broad abilities and nearly 100 narrow abilities that have been identified for humans (Schneider & McGrew, 2012, in press). Yet even that large number is probably an underestimate of human capabilities because certain aspects of human intelligence have been seriously overlooked in IQ tests (social intelligence) or trivialized (creativity, see Kaufman, Kaufman, & Lichtenberger, 2011).

The one main exception to ignoring social intelligence in human tests is the inclusion of a Theory of Mind subtest on the NEPSY-II (a developmental NEuroPSYcological asssessment; Korkman, Kirk, & Kemp, 2007), but that test is considered a neuropsychological battery, not an IQ test. Even though the WPPSI-IV and other Wechsler scales include the Comprehension subtest, intended to assess social judgment and reasoning, a principal difference remains between the PCTB and all human IQ tests—the former embraces the measurement of social cognition whereas the latter virtually ignores it. As a result, any comparison between the PCTB and human IQ tests is not an exact one. The abilities called "social" by the PCTB are generally characterized as Gc by CHC theory. Gc is traditionally described as a more factual type of knowledge, the type of information learned in school and retained onward. The more formal definition, however, encompasses under the umbrella of "culture" many abilities characterized as social or cultural by the PCTB—theory of mind, gestural communication, gaze following, and imitation, for example. These aspects of social learning are learned informally by humans and nonhuman animals but still are aspects of Gc:

> Comprehension-knowledge (Gc): The knowledge of the culture that is incorporated by individuals vis-à-vis a process of acculturation. Gc is typically described as a person's breadth and depth of acquired knowledge of the language, information and concepts of a specific culture, and/or the application of this knowledge. Gc is primarily a store of verbal or language-based declarative (knowing what) and procedural (knowing how) knowledge acquired through the investment of other abilities during formal and informal educational and general life experiences. (McGrew, 2009, p. 11)

As a result, when applying CHC theory to a test such as the PCTB, it is important to keep in mind that although these social abilities may be included in Gc in theory, they may not always be considered in practical application, nor in creation of assessments or in their interpretation. *Gc* currently includes six narrow abilities: General (verbal) Information, Language Development, Lexical Knowledge, Listening Ability, Communication Ability, and Grammatical Sensitivity. With the exception of grammatical sensitivity, these are good matches for the abilities labeled as the social component by the PCTB—attention, production, comprehension, and gaze following—and hypothesized to develop exponentially in children after age 2½ years.

The focus on types of intelligence also relates to recent evidence showing superior Gc in social living species of orangutans, as opposed to solitary living species raised in the same circumstances (Forss, Willems, Call, & van Schaik, 2016); in "enculturated" chimpanzees, as opposed to ones reared in laboratory settings (Russell et al., 2011); in dogs, which live cooperatively with humans, as opposed to wolves (Miklósi, Topál, & Csányi, 2004); and even in a population of foxes experimentally bred for compatibility with people (Trut, 1999). Hare (2017) summarized this idea and expanded on the cultural intelligence hypothesis with his human self-domestication hypothesis, which asserts that in the more recent stages of both human and nonhuman animal evolution, sociability and cultural understanding have become far more adaptive than aggression. He cited as an example a population of foxes bred for tame relationships with humans that also showed increasingly early auditory responses relative to wild foxes, eyes that opened earlier than their counterparts, and fear responses that developed later. None of these traits were selected for, and they were all noted to be on a spectrum between wild foxes and domestic dogs, meaning wild foxes showed the earliest fear response, domestic foxes' response was slightly delayed, and domestic dogs had the latest developing fear response (Trut, 1999). Increases in these traits also seem conducive to better performance in many of the narrow abilities considered to be part of crystalized intelligence.

It will be interesting to compare factor structure for apes reared by their mothers and those reared in "nursery" situations (infant groups with basic care provided by humans; i.e., animals who had been removed from their mothers for various reasons), which would have had slightly more exposure to human cultural practices. Russell et al. (2011) found no difference between these two groups, although they found both groups to be less successful at the social cognition portion of the PCTB than apes raised in environments categorized as "encultured" (i.e., language research centers such as the one at Great Ape Trust of Iowa and the Language Research Center in Atlanta Georgia; Russell et al., 2011). Also performing better on social tasks were chimpanzees who did better on measures of imitation (Pope, Russell, & Hopkins, 2015), a skill linked to the development of cultural traditions (Hopper et al., 2007).

Limitations

Maximum likelihood estimation in CFA is based on large sample theory. Although most models estimated in this study were not complex, small sample sizes are related to less trustworthy results. One limitation of this study that needs to be considered is the smaller sample size—especially when considered in relation to CFAs with human intelligence data.

Test–retest stability measured for normal children on IQ tests is conducted over a matter of weeks or months, whereas the data from the Hopkins chimpanzees were over a 2½–3-year interval, so the comparison of stability correlations may not be fair. Data from mandatory special education evaluations in the United States provide a better comparison with data on the chimpanzees. Test–retest data over a 3-year period in humans often show notably less stability over time for children from special populations, such as learning disabilities, behavior disorders, and intellectual disabilities, with coefficients as low as .58 for Full Scale IQ on the WISC–R (Wechsler Intelligence Scale for Children - Revised; Anderson, Cronin, & Kazmierski, 1989) and .56 for subtest median on the WISC-IV (Wechsler Intelligence Scale for Children 4; Watkins & Smith, 2013). The value in the high .50s for IQ is not substantially larger than the value of .50 for PCTB composite score for chimpanzees in the present study. The low stability of the PCTB tests (median $r = .24$), however, should be investigated more—further it would be interesting to see what the stability is like over a shorter period of time.

Conclusions

The PCTB has a good Gv factor; not as well defined, but an acceptable Gc factor; and a *g* factor. All are invariant and stable when the test is administered to chimpanzees over time. The performances on specific tasks that compose the PCTB, however, were *not* stable over time. There is some evidence of a Gf factor, but perhaps one that splits in two, and a Glr factor, but this evidence is weaker. There appears to be some overlap in the description of ape abilities and abilities from human CHC theory, and although that connection needs to be described better, measures for apes and humans seem to be evaluating similar constructs in each species. However, our best CFA models did not fully generalize across time or across samples, indicating that further cross-validation is needed.

References

Anderson, B. (1993). Evidence from the rat for a general factor that underlies cognitive performance and that relates to brain size: Intelligence? *Neuroscience Letters, 153,* 98–102. http://dx.doi.org/10.1016/0304-3940(93)90086-Z

Anderson, P. L., Cronin, M. E., & Kazmierski, S. (1989). WISC-R stability and re-evaluation of learning-disabled students. *Journal of Clinical Psychology, 45,* 941–944. http://dx.doi.org/10.1002/1097-4679(198911)45:6<941::AID-JCLP2270450619>3.0.CO;2-P

Arden, R., & Adams, M. J. (2016). A general intelligence factor in dogs. *Intelligence, 55,* 79–85. http://dx.doi.org/10.1016/j.intell.2016.01.008

Balda, R. P., Kamil, A. C., & Bednekoff, P. A. (1996). Predicting cognitive capacity from natural history: Examples from four species of corvids. *Current Ornithology, 13,* 33–66. http://dx.doi.org/10.1007/978-1-4615-5881-1_2

Binet, A., & Simon, T. (1904). Méthodes nouvelles pour le diagnostic du niveau intellectual des anormaux [New methods for the diagnosis of the intellectual level of the abnormal]. *L'Année Psychologique, 11,* 191–244. http://dx.doi.org/10.3406/psy.1904.3675

Bond, A. B., Kamil, A. C., & Balda, R. P. (2003). Social complexity and transitive inference in corvids. *Animal Behaviour, 65,* 479–487. http://dx.doi.org/10.1006/anbe.2003.2101

Boogert, N. J., Anderson, R. C., Peters, S., Searcy, W. A., & Nowicki, S. (2011). Song repertoire size in male song sparrows correlates with detour reaching, but not with other cognitive measures. *Animal Behaviour, 81,* 1209–1216. http://dx.doi.org/10.1016/j.anbehav.2011.03.004

Burgess, G. C., & Braver, T. S. (2010). Neural mechanisms of interference control in working memory: Effects of interference expectancy and fluid intelligence. *PLoS ONE, 5,* e12861. http://dx.doi.org/10.1371/journal.pone.0012861

Burkart, J. M., Schubiger, M. N., & Van Schaik, C. P. (2016). The evolution of general intelligence. *Behavioral and Brain Sciences, 40,* e195. http://dx.doi.org/10.1017/S0140525X16000959

Canivez, G. L. (2013). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwean (Eds.), *The Oxford handbook of child psychological assessments* (pp. 84–112). New York, NY: Oxford University Press.

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children, Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment, 29,* 458–472. http://dx.doi.org/10.1037/pas0000358

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511571312

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin, 38,* 592.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54,* 1–22. http://dx.doi.org/10.1037/h0046743

Deaner, R. O., van Schaik, C. P., & Johnson, V. (2006). Do some taxa have better domain-general cognition than others? A meta-analysis of nonhuman primate studies. *Evolutionary Psychology*. Advance online publication. http://dx.doi.org/10.1177/147470490600400111

Flanagan, D. P., Ortiz, S., & Alfonso, V. (2013). *Essentials of cross-battery assessment* (3rd ed.). New York, NY: Wiley.

Forss, S. I. F., Willems, E., Call, J., & van Schaik, C. P. (2016). Cognitive differences between orang-utan species: A test of the cultural intelligence hypothesis. *Scientific Reports, 6,* 30516. http://dx.doi.org/10.1038/srep30516

Galsworthy, M. J., Paya-Cano, J. L., Liu, L., Monleón, S., Gregoryan, G., Fernandes, C., . . . Plomin, R. (2005). Assessing reliability, heritability and general cognitive ability in a battery of cognitive tasks for laboratory mice. *Behavior Genetics, 35,* 675–692. http://dx.doi.org/10.1007/s10519-005-3423-9

Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences*. London, United Kingdom: Macmillan. http://dx.doi.org/10.1037/13474-000

Galton, F. (1883). *Inquiries into human faculty and its development*. London, United Kingdom: Macmillan. http://dx.doi.org/10.1037/14178-000

Hare, B. (2017). Survival of the friendliest: *Homo sapiens* evolved via selection for prosociality. *Annual Review of Psychology, 68,* 155–186. http://dx.doi.org/10.1146/annurev-psych-010416-044201

Herrmann, E., & Call, J. (2012). Are there geniuses among the apes? *Philosophical Transactions of the Royal Society B: Biological Sciences, 367,* 2753–2761. http://dx.doi.org/10.1098/rstb.2012.0191

Herrmann, E., Call, J., Hernàndez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317,* 1360–1366. http://dx.doi.org/10.1126/science.1146282

Herrmann, E., Hernández-Lloreda, M. V., Call, J., Hare, B., & Tomasello, M. (2010). The structure of individual differences in the cognitive abilities of children and chimpanzees. *Psychological Science, 21,* 102–110. http://dx.doi.org/10.1177/0956797609356511

Herrnstein, R. J., & Murray, C. A. (1994). *The bell curve: Intelligence and class structure in American life*. New York, NY: Free Press.

Hopkins, W. D., Russell, J. L., & Schaeffer, J. (2014). Chimpanzee intelligence is heritable. *Current Biology, 24,* 1649–1652. http://dx.doi.org/10.1016/j.cub.2014.05.076

Hopper, L. M., Spiteri, A., Lambeth, S. P., Schapiro, S. J., Horner, V., & Whiten, A. (2007). Experimental studies of traditions and underlying transmission processes in chimpanzees. *Animal Behaviour, 73,* 1021–1032. http://dx.doi.org/10.1016/j.anbehav.2006.07.016

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review, 75,* 242–259. http://dx.doi.org/10.1037/h0025662

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57,* 253–270. http://dx.doi.org/10.1037/h0023816

Hornung, C., Brunner, M., Reuter, R. A. P., & Martin, R. (2011). Children's working memory: Its structure and relationship to fluid intelligence. *Intelligence, 39,* 210–221. http://dx.doi.org/10.1016/j.intell.2011.03.002

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. http://dx.doi.org/10.1080/10705519909540118

Isden, J., Panayi, C., Dingle, C., & Madden, J. (2013). Performance in cognitive and problem-solving tasks in male spotted bowerbirds does not correlate with mating success. *Animal Behaviour, 86,* 829–838. http://dx.doi.org/10.1016/j.anbehav.2013.07.024

Jensen, A. R. (1998). *The science of mental ability*. Westport, CT: Praeger.

Kaufman, A. S. (2009). *IQ Testing 101*. New York, NY: Springer.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children* (A. G. Services, Ed.). Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children–Second Edition (KABC-II)*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2018). *Kaufman Assessment Battery for Children–Second Edition/Normative Update (KABC-II NU)*. Bloomington, MN: Pearson.

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016). *Intelligent testing with the WISC-V*. Hoboken, NJ: Wiley.

Kaufman, A. B., Reynolds, M. R., & Kaufman, A. S. (2013). *Psychometric Analysis of the abilities of apes and 2½ year-olds*. Honolulu, HI: American Psychological Association. http://dx.doi.org/10.1037/e604992013-001

Kaufman, J. C., Kaufman, S. B., & Lichtenberger, E. O. (2011). Finding creativity on intelligence tests via divergent production. *Canadian Journal of School Psychology, 26,* 83–106. http://dx.doi.org/10.1177/0829573511406511

Keith, T. Z. (2015). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling*. New York, NY: Routledge.

Keith, T. Z., & Reynolds, M. R. (2012). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In P. L. Harrison & D. P. Flanagan (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 758–799). New York, NY: Guilford Press.

Kolata, S., Light, K., & Matzel, L. D. (2008). Domain-specific and domain-general learning factors are expressed in genetically heterogeneous CD-1 mice. *Intelligence, 36,* 619–629. http://dx.doi.org/10.1016/j.intell.2007.12.001

Kolata, S., Light, K., Townsend, D. A., Hale, G., Grossman, H. C., & Matzel, L. D. (2005). Variations in working memory capacity predict individual differences in general learning abilities among genetically diverse mice. *Neurobiology of Learning and Memory, 84,* 241–246. http://dx.doi.org/10.1016/j.nlm.2005.07.006

Korkman, M., Kirk, U., & Kemp, S. L. (2007). *NEPSY II. Clinical and interpretative manual*. San Antonio, TX: Psychological Corporation.

Lacreuse, A., Russell, J. L., Hopkins, W. D., & Herndon, J. G. (2014). Cognitive and motor aging in female chimpanzees. *Neurobiology of Aging, 35,* 623–632. http://dx.doi.org/10.1016/j.neurobiolaging.2013.08.036

Locurto, C., & Scanlon, C. (1998). Individual differences and a spatial learning factor in two strains of mice (*Mus musculus*). *Journal of Comparative Psychology, 112,* 344–352. http://dx.doi.org/10.1037/0735-7036.112.4.344

MacLean, E. L., Herrmann, E., Suchindran, S., & Hare, B. (2017). Individual differences in cooperative communicative skills are more similar between dogs and humans than chimpanzees. *Animal Behaviour, 126,* 41–51. http://dx.doi.org/10.1016/j.anbehav.2017.01.005

MacLean, E. L., Matthews, L. J., Hare, B. A., Nunn, C. L., Anderson, R. C., Aureli, F., . . . Wobber, V. (2012). How does cognition evolve? Phylogenetic comparative psychology. *Animal Cognition, 15,* 223–238. http://dx.doi.org/10.1007/s10071-011-0448-8

Matzel, L. D., Han, Y. R., Grossman, H., Karnik, M. S., Patel, D., Scott, N., . . . Gandhi, C. C. (2003). Individual differences in the expression of a “general” learning ability in mice. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 23,* 6423–6433. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12878682

Matzel, L. D., & Sauce, B. (2017). Individual differences: Case studies of rodent and primate intelligence. *Journal of Experimental Psychology: Animal Learning and Cognition, 43,* 325–340. http://dx.doi.org/10.1037/xan0000152

McGrew, K. (2009). *Cattell-Horn-Carroll (CHC) broad and narrow cognitive ability definitions*. Retrieved from http://www.iapsych.com/articles/chcdefs031109.pdf

Miklósi, A., Topál, J., & Csányi, V. (2004). Comparative social cognition: What can dogs teach us? *Animal Behaviour, 67,* 995–1004. http://dx.doi.org/10.1016/j.anbehav.2003.10.008

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105–125. http://dx.doi.org/10.1037/1082-989X.7.1.105

Niileksela, C. R., Reynolds, M. R., & Kaufman, A. S. (2013). An alternative Cattell-Horn-Carroll (CHC) factor structure of the WAIS-IV: Age invariance of an alternative model for ages 70–90. *Psychological Assessment, 25,* 391–404. http://dx.doi.org/10.1037/a0031175

Piaget, J. (1936). *Origins of intelligence in the child*. London, United Kingdom: Routledge & Kegan Paul.

Plomin, R. (2001). The genetics of g in human and mouse. *Nature Reviews Neuroscience, 2,* 136–141. http://dx.doi.org/10.1038/35053584

Pope, S. M., Russell, J. L., & Hopkins, W. D. (2015). The association between imitation recognition and socio-communicative competencies in chimpanzees (*Pan troglodytes*). *Frontiers in Psychology, 6,* 188. http://dx.doi.org/10.3389/fpsyg.2015.00188

Reader, S. M., Hager, Y., & Laland, K. N. (2011). The evolution of primate general and cultural intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366,* 1017–1027. http://dx.doi.org/10.1098/rstb.2010.0342

Reise, S. P. (2012). Invited paper: The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47,* 667–696. http://dx.doi.org/10.1080/00273171.2012.715555

Reynolds, M. R., Keith, T. Z., Flanagan, D. P., & Alfonso, V. C. (2013). A cross-battery, reference variable, confirmatory factor analytic investigation of the CHC taxonomy. *Journal of School Psychology, 51,* 535–555. http://dx.doi.org/10.1016/j.jsp.2013.02.003

Rushton, J. P. (1997). Race, intelligence, and the brain. *Personality and Individual Differences, 23,* 169–189. http://dx.doi.org/10.1016/S0191-8869(97)80984-1

Russell, J. L., Lyn, H., Schaeffer, J. A., & Hopkins, W. D. (2011). The role of socio-communicative rearing environments in the development of social and physical cognition in apes. *Developmental Science, 14,* 1459–1470. http://dx.doi.org/10.1111/j.1467-7687.2011.01090.x

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research, 8,* 23–74. Retrieved from http://psycnet.apa.org/record/2003-08119-003

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll (CHC) model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York, NY: Guilford Press.

Schneider, W. J., & McGrew, K. S. (in press). The Cattell-Horn-Carroll (CHC) model of intelligence. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed.). New York, NY: Guilford Press.

Schrank, F. A., Dekker, S. L., & Garruto, J. M. (2016). *Essentials of WJ IV Cognitive Abilities assessment*. Hoboken, NJ: Wiley.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside.

Shaw, R. C., Boogert, N. J., Clayton, N. S., & Burns, K. C. (2015). Wild psychometrics: Evidence for “general” cognitive performance in wild New Zealand robins, *Petroica longipes*. *Animal Behaviour, 109,* 101–111. http://dx.doi.org/10.1016/j.anbehav.2015.08.001

Spearman, C. (1904). “General Intelligence,” objectively determined and measured. *The American Journal of Psychology, 15,* 201–292. http://dx.doi.org/10.2307/1412107

Spearman, C. (1927). *The abilities of man*. New York, NY: MacMillan and Company.

Tomasello, M., & Herrmann, E. (2010). Ape and human cognition: What’s the difference? *Current Directions in Psychological Science, 19,* 3–8. http://dx.doi.org/10.1177/0963721409359300

Trut, L. N. (1999). Early canid domestication: The farm-fox experiment. *American Scientist, 87,* 160–169. http://dx.doi.org/10.1511/1999.2.160

Vonk, J., & Povinelli, D. (2011). Individual differences in long-term cognitive testing in a group of captive chimpanzees. *International Journal of Comparative Psychology, 24,* 137–167.

Watkins, M. W., & Smith, L. G. (2013). Long-term stability of the Wechsler Intelligence Scale for Children, Fourth Edition. *Psychological Assessment, 25,* 477–483. http://dx.doi.org/10.1037/a0031653

Wechsler, D. (1939). *Measurement of adult intelligence*. Baltimore, MD: Williams & Wilkins. http://dx.doi.org/10.1037/10020-000

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence* (4th ed.). San Antonio, TX: Pearson.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.

Woodley of Menie, M. A., Fernandes, H. B. F., & Hopkins, W. D. (2015). The more *g*-loaded, the more heritable, evolvable, and phenotypically variable: Homology with humans in chimpanzee cognitive abilities. *Intelligence, 50,* 159–163. http://dx.doi.org/10.1016/j.intell.2015.04.002

Yoakum, C. S., & Yerkes, R. M. (Eds.). (1920). *Army mental tests*. New York, NY: Henry Holt. http://dx.doi.org/10.1037/11054-000

Received August 30, 2017
Revision received May 1, 2018
Accepted May 21, 2018 ■