Intelligence 50 (2015) 87–92

Contents lists available at ScienceDirect

Intelligence

Spearman's hypothesis and Amerindians: A meta-analysis

Jan te Nijenhuis[a,*], Michael van den Hoek[a], Elijah L. Armstrong[b]

[a] *Work and Organizational Psychology, University of Amsterdam, The Netherlands*
[b] *Fairfax, CA 94390, United States*

ARTICLE INFO

Article history:
Received 14 November 2014
Received in revised form 16 January 2015
Accepted 20 February 2015
Available online xxxx

Keywords:
Spearman's hypothesis
Amerindians
IQ
g loadings
Group differences

ABSTRACT

Spearman's hypothesis states that differences between groups on the subtests of an IQ battery are a function of the *g* loadings of these subtests, such that there are small differences between groups on subtests with low *g* loadings and large differences between groups on subtests with high *g* loadings, and it is confirmed in the large majority of studies. In this paper, we test Spearman's hypothesis, comparing Amerindians with Whites in the US and Canada. We carried out a meta-analysis based on 25 data points and a total $N = 2706$ Amerindians. Spearman's hypothesis was strongly confirmed with a mean r with a value of .62. We conclude that Spearman's hypothesis appears to be a more regular phenomenon than previously thought.

1. Introduction

Group differences in IQ are common and often quite large. For example, Blacks in the US have IQs about one standard deviation below their White counterparts (e.g., Jensen, 1998; Lynn, 2006; Rushton & Jensen, 2005). The IQ of Amerindians is approximately 86 (Lynn, 2006), and that of Ashkenazi Jews is approximately 110 (Lynn, 2011).

Nearly all mental tests correlate to some extent (Jensen, 1969, 1998; Spearman, 1927). This means that nearly all tests a share common variance, and the cross-test common variance is called *g*, or 'general intelligence'. When a test is a good measure of *g* (i.e., it predicts the results of other tests better, or to use a more technical definition, has a stronger loading on the first principal component extracted from different correlated tests), it is said to have a high *g* loading. Analogously, when a test poorly measures *g*, its *g* loading is low. One interesting finding about group IQ differences is that they tend to be more pronounced on more *g*-loaded subtests (i.e., subtests that are better measures of the first principal component in a correlation matrix of tests; Jensen, 1998). This pattern has been found for the large majority of group comparisons: for example, Whites and Blacks in the US (Jensen, 1998), between European Whites and immigrants to Western Europe (te Nijenhuis & van der Flier, 1997, 2003), and between US Whites and Africans (Rushton & Jensen, 2003). However, several papers show that Spearman's hypothesis is not always confirmed, such as for White and Black prisoners (Jensen & Faulstich, 1988), Whites and Asians (Kane, 2007), and the Sámi in Lapland (Armstrong, Woodley, & Lynn, 2014). The *g*-loadedness of racial IQ gaps is referred to as *Spearman's hypothesis* (e.g., Jensen, 1985).

The Amerindians are the group descended from the autochthons of both American continents. This group is physically and genetically differentiated from other population groups (Cavalli-Sforza, Menozzi, & Piazza, 1994; Lynn, 2006). The Amerindian population is a relatively small minority in North America and only makes up about 0.75% of the US and 5.6% of the Canadian population (US Census Bureau, 2007; Statistics Canada, 2011). While this is a relatively small population, there is a fair amount of intelligence research on Amerindians. Generally this research shows that Amerindians score significantly lower on intelligence tests than Whites. Their profile is that of higher Performance scores than Verbal scores.

* Corresponding author at: Gouden Leeuw 746, 1103 KR Amsterdam, The Netherlands.
E-mail address: JanteNijenhuis@planet.nl (J. te Nijenhuis).

http://dx.doi.org/10.1016/j.intell.2015.02.006
0160-2896/

Is the Amerindian/White IQ gap *g* loaded or not? The amount of empirical studies is extremely modest: only one study was carried out and it found support for Spearman's hypothesis (Reynolds, Willson, & Ramsey, 1999). In the present study, we evaluate Spearman's hypothesis for Amerindians using a large array of studies.

1.1. Research questions

We carried out a meta-analysis so we could estimate the more reliable meta-analytical correlation between White–Amerindian group differences and *g* loadings. This will answer the question whether Spearman's hypothesis can be generalized to Amerindians or whether the group of Amerindians functions as an outlier.

2. Method

2.1. Meta-analysis

The purpose of this study is to determine whether the correlation between the magnitude of *g* loadings and difference scores on the IQ subtest scores between Whites and Amerindians is strongly positive in sign. We carried out a meta-analysis where we tested Spearman's hypothesis with Whites and Amerindians in the US and Canada. Meta-analytical echniques (Hunter & Schmidt, 2004) were applied using the software package developed by Schmidt and Le (2004). We carried out bare-bones meta-analyses where we corrected for only one artifact, namely sampling error.

2.2. Inclusion rules

For studies to be included in a meta-analysis two criteria had to be met: First, in order to obtain a reliable estimate of the true correlation between each of the variables and *g* loadings, the cognitive batteries had to be based on a minimum of five subtests. Second, the IQ test had to be well-validated.

2.3. Searching and screening studies

Digital and manual searches were carried out in a search for data. Data from the US Census Bureau (2012) was used to identify Amerindian tribes for use in searching for studies. The following tribes were identified using these data: Aleut, Apache, Arapaho, Athabasca, Blackfeet, Cherokee, Cheyenne, Chickasaw, Chippewa, Choctaw, Colville, Comanche, Cree, Creek, Crow, Delaware, Hopi, Houma, Inupiat, Iroquois, Kiowa, Lumbee, Menominee, Navajo, Osage, Ottawa, Paiute, Pima, Potawatomi, Pueblo, Salish, Seminole, Shoshone, Tlingit-Haida, Tohono O'Odham, Tsimshian, Ute, Yakama, Yaqui, Yuman, and Yup'ik. For the digital search we used several electronic databases: Google Scholar, ProQuest, PsycINFO, and Catalogue Plus (Primo). All the tribe names were used as keywords along with more general terms like 'Native American', 'Amerindian', and 'Indians'. These were combined with the keywords 'intelligence', 'IQ', 'mental ability', 'mental capacity', 'cognitive ability', 'aptitude', 'competence', 'differences', 'WISC', 'WAIS', 'WPPSI', 'K-ABC', and 'Woodcock Johnson'.

Manual searches were also carried out; the journal *Psychology in Schools* contained many studies on the intelligence of Amerindians and all editions available at the University of Amsterdam were manually checked for usable studies. Reference lists of all currently included empirical studies were checked to identify any potential articles that may have been missed by earlier search methods. Finally, several researchers who have conducted research on the mental abilities of Indians were contacted in order to obtain any additional articles or Supplementary information. This led to contacts with the Department of Special Collections at the University of North Dakota which supplied us with papers from the Damian Vraniak archive (Vraniak was a prolific researcher in this field).

We ended up with 13 studies containing 25 data points. The studies by Newman et al. (2009) and McShane (1980) used composite scores, so they were not comparable to the other studies that used subtests, and therefore they were not included in the meta-analysis. We would like to note that, although there are many studies reporting IQ scores of Amerindians, studies reporting scores on all subtests of an IQ battery are quite rare.

2.4. g loadings

To test Spearman's hypothesis Pearson correlations between *d* scores of the subtests of the IQ battery and *g* loadings were computed. In general, *g* loadings were computed by conducting a principal-axis factor analysis on the correlation matrix of a test battery's subtest scores. The subtest's loadings on the first unrotated factor indicate the subtest's loading on *g*. *g* loadings were always matched to the age range of the groups involved in the comparison as closely as possible. If the age range of the comparison groups comprised more than one age group of the IQ battery, we computed weighted average *g* loadings of all the age groups of the IQ battery that fall within the age range of the comparison groups.

2.5. Computation of score differences

To compute the score differences between a White/norm group and an Amerindian group (*d*), the mean score of the Amerindian group was subtracted from the mean score of the US standardization sample. The difference was divided by the *SD* of the standardization sample.

2.6. Computation of correlation between g and d

We computed the Pearson correlations between *g* loadings and group differences for all studies. To check whether the profile of higher Performance scores and lower Verbal scores influenced the outcomes we also compute $r(g \times d)$ for, respectively, all Performance subtests, and all Verbal subtests.

3. Results

The results of the studies on the vector correlation between *g* loadings and the score differences between Amerindians and Whites (*d*) are shown in Table 1. The Table shows data derived from 13 studies, yielding 25 data correlations, with participants numbering a total of 24, 830 Whites, 2706 Amerindians, and with a total harmonic $N = 9266$. It also lists the reference for the study, the Amerindian tribe, the cognitive ability test used,

Table 1
Studies of correlations between g loadings and Amerindian/White differences using all subtests.

Reference	Tribe	Test	r ($d \times g$)	$N_{subtests}$	N_{White}	$N_{Native\text{-}American}$	$N_{harmonic}$	Age (range)
Turner and Penfold (1952)	Chippewa, Muncey, and Oneida	WISC	.88	10	200	31	107	7
		WISC	.64	10	200	25	89	10
		WISC	.70	10	200	26	92	14
Howell, Evans, and Downing (1958)	Navajo	WAIS	.69	11	200	100	267	16.5[a] (16–17)
Cundick (1970)	Unknown	WPPSI	.72	11	600	27	103	4[a] (3–5)[a]
		WISC	.87	12	200	26	92	7[a] (6–8)[a]
John, Krichev, and Bauman (1976)	Ojibwa and Cree	WISC	.85	10	200	33	113	6.5[a] (6–7)
		WISC	.81	10	200	31	107	9.5[a] (9–10)
		WISC	.73	10	200	36	122	14.5[a] (14–15)
Teeter, Moore, and Petersen (1982)	Navajo							
Non-handicapped		WISC-R	.51	11	2200	113	430	11[a] (6–16)
Educ. disadvantaged[b]		WISC-R	.48	11	2200	189	696	
Learning disabled[b]		WISC-R	.46	11	2200	150	562	
Connelly (1983) [b]	Tlingit	WISC-R	.10	11	1000	61	230	8[a] (6–10)
		WISC-R	.39	11	1200	85	318	13.5[a] (11–16)
Mitchell (1985)[b]	Cherokee and Kiowa	K-ABC	.54	8	42	44	86	8.5[a] (6–11)
		WISC-R	.54	10	1200	44	170	
Snyder (1991)[b]	Unknown	WISC-R	.15	11	1800	48	187	10.5[a] (6.5–14.5)
Reynolds et al. (1999)	Papagos	WISC-R	.69	12	2200	240	866	10.12 (6–16)
Kush and Watkins (2007)[b]	12 different tribes	WISC-III	.58	12	1543	334	1098	10 (6–16)
Vanderpool and Catano (2008)								
Average education	First Nation	Various	.25	6	108	101	209	31.5[a] (18–45)
Low education	First Nation	Various	.25	6	108	51	139	
Smith (2009)	Crow	WISC-III	.63	11	1543	205	724	10.52 (6–16)
	Cheyenne	WISC-III	.60	11	1543	292	982	
	Sioux	WISC-III	.61	11	1543	238	825	
Nakano and Watkins (2013)[b]	40% Navajo, 60% unknown	WISC-IV	.53	10	2200	176	652	10.6 (6–16)

Note: $N_{harmonic}$ is computed using the formula $\frac{4}{\frac{1}{n1}+\frac{1}{n2}}$ where $n1$ and $n2$ are the amounts of participants in groups $n1$ and $n2$, respectively.
[a] Estimated.
[b] Referral sample or mostly a referral sample or special education sample.

the vector correlation between g loadings and d, and the sample size. The correlations are positive in sign and the large majority of them are substantial in magnitude. Table 2 presents the results of the bare-bone meta-analysis of the 25 data points. Table 2 shows the number of correlation coefficients (K), total sample size (N), the mean-weighted vector correlation (mean r), and the standard deviation of the vector correlation (SD_r). The last column presents the percentage of variance explained by a sampling error (%VE). The analysis of all 25 data points yields a mean vector correlation of .56, with 5.8% of the variance in the observed correlations explained by the sampling error. This percentage is very low and suggests the presence of a strong moderator or several moderators.

Table 2
Exploratory bare-bones meta-analytical results for correlations between g loadings and Amerindian/White differences.

Studies included	K	N_h	Mean r	SD_r	%VE
All studies	25	9266	.56	.149	5.8
Most-representative studies	16	5267	.62	.127	7.1
Most-representative studies minus two outliers	14	4919	.65	.082	14.2
Less-representative studies	9	3999	.47	.132	7.8

Note. Bare-bones meta-analytical results: Score differences between Amerindians and Whites, and g loadings. K = number of correlations; N = total sample size; mean r = mean-weighted vector correlation; SD_r = standard deviation of observed correlation; %VE = percentage of variance accounted for by sampling errors.

We tested whether the quality of the data points acted as a moderator variable. Some of the samples were of lesser quality, because they consisted of children that were referred to psychoeducational testing to determine eligibility for special education, or they consisted of children in special education schools. In Table 2 we also present the meta-analytical findings without these samples. Fig. 1 shows a scatter plot of these 16 data points. When excluding the aforementioned studies from

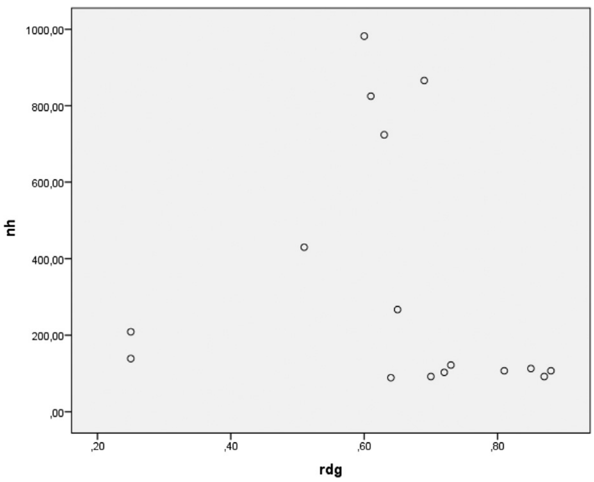

Fig. 1. Scatter plot of r ($d \times g$) and $N_{harmonic}$ for Amerindian studies with less-representative samples removed.

the analysis, the mean vector correlation rises to .62, whereas the mean vector correlation is .47 for the nine less-representative studies. So, quality of the data points is a clear moderator variable.

The percentage variance explained by the sampling error in the 16 most-representative studies is only a very modest 7.1, which is still suggestive of the presence of strong moderator variables. We tested whether the number of subtests in the various IQ batteries acted as a moderator variable. The values of .25 and .25, respectively, for the two data points from Vanderpool and Catano (2008) are both based on only six subtests, most likely making it the least reliable test of Spearman's hypothesis in this meta-analysis, which could explain the low value of the correlation. In a final analysis we left out these two outliers, and Table 2 shows that then the mean vector correlation rises to .65 with now 14.2% of the variance explained in the 14 data points by the sampling error. However, this is only a very small increase in the size of the correlation, and also a very modest increase in the amount of variance explained. Moreover, a test of Spearman's hypothesis based on six subtests can still be considered adequate. So, the value of .62 is the best estimate of the meta-analytical correlation between intelligence and group differences.

Table 3 also shows the results when Spearman's hypothesis is tested on, respectively, all Verbal, and all Performance subtests. Table 4 shows that the meta-analytical r for all Verbal subtests has a value of .73, which is comparable to the value of the meta-analytical r when testing Spearman's hypothesis on all the subtests. However, Table 5 shows that the meta-analytical r for all Performance subtests is −.21, so showing a small anti-Jensen effect.

Table 4
Bare-bones meta-analytical results for correlations between g loadings and Amerindian/White differences on verbal subtests (minus Mitchell K-ABC).

Studies included	K	N_h	Mean r	SD_r	%VE
All studies	24	9180	.73	.326	0.6
Most-representative studies	16	5267	.80	.234	0.7
Less-representative studies	8	3913	.63	.398	0.5

Note. Bare-bones meta-analytical results: Score differences between Amerindians and Whites, and g loadings. K = number of correlations; N = total sample size; mean r = mean-weighted vector correlation; SD_r = standard deviation of observed correlation; and %VE = percentage of variance accounted for by sampling errors.

4. Conclusion

The goal of this study was to explore whether differences in the IQ profile between Amerindians and Whites have a strong correlation with general intelligence, and we obtained a meta-analytic vector correlation of .62. As these findings are based on the complete world literature on this topic and a large total N = 2706 Amerindians, it is possible to draw strong conclusions, namely that group differences in intelligence between Amerindians and Whites are strongly related to general intelligence. Spearman's hypothesis appears to be a more regular phenomenon than previously thought (Table 3).

The sample size varied between the studies in the meta-analysis, and the sampling error explained only a very modest amount of variance between the data points. This means that there are still moderators left, or that other statistical artifacts than the sampling error (Hunter & Schmidt, 2004) explain the rest of the variation in the data point. Still, the meta-analytical

Table 3
Studies of correlations between g loadings and Amerindian/White differences.

Reference	Tribe	Test	r ($d \times g$)	r_{Verb}	r_{Perf}	$N_{subtests}$	N_{White}	$N_{Native\text{-}American}$	$N_{harmonic}$
Turner and Penfold (1952)	Chippewa, Muncey, and Oneida	WISC	.88	.85	.64	10	200	31	107
		WISC	.64	.89	.10	10	200	25	89
		WISC	.70	.72	−.31	10	200	26	92
Howell et al. (1958)	Navajo	WAIS	.69	.73	.39	11	200	100	267
Cundick (1970)	Unknown	WPPSI	.72	−.22	.45	11	600	27	103
		WISC	.87	.86	.67	12	200	26	92
John et al. (1976)	Ojibwa and Cree	WISC	.85	.68	.37	10	200	33	113
		WISC	.81	.32	.20	10	200	31	107
		WISC	.73	−.17	.02	10	200	36	122
Teeter et al. (1982)	Navajo								
Non-handicapped		WISC-R	.51	.84	−.49	11	2200	113	430
Educ. disadvantaged[2]		WISC-R	.48	.88	−.49	11	2200	189	696
Learning disabled[2]		WISC-R	.46	.80	−.40	11	2200	150	562
Connelly (1983)[2]	Tlingit	WISC-R	.10	−.50	−.41	11	1000	61	230
		WISC-R	.39	.55	−.52	11	1200	85	318
Mitchell (1985)[2]	Cherokee and Kiowa	K-ABC	.54	–	.55	8	42	44	86
		WISC-R	.54	.04	−.74	10	1200	44	170
Snyder (1991)[2]	Unknown	WISC-R	.15	−.28	−.58	11	1800	48	187
Reynolds et al. (1999)	Papagos	WISC-R	.69	.92	.41	12	2200	240	866
Kush and Watkins (2007)[2]	12 different tribes	WISC-III	.58	.84	−.18	12	1543	334	1098
Vanderpool and Catano (2008)									
Average education	First Nation	Various	.25	.67	–	6	108	101	209
Low education	First Nation	Various	.25	.72	–	6	108	51	139
Smith (2009)	Crow	WISC-III	.63	.85	−.22	11	1543	205	724
	CheyenneSioux	WISC-III	.60	.91	−.32	11	1543	292	982
		WISC-III	.61	.88	−.22	11	1543	238	825
Nakano and Watkins (2013)[2]	40% Navajo, 60% unknown	WISC-IV	.53	.70	−.67	10	2200	176	652

Note. The K-ABC data from Mitchell (1985) did not have enough verbal subtests to calculate verbal r_{dg}. The study from Vanderpool & Canto (2008) did not have enough non-verbal tests to calculate performance r_{dg}.

[2] Referral sample or mostly a referral sample or special education sample.

Table 5
Bare-bones meta-analytical results for correlations between *g* loadings and Amerindian/White differences on performance subtests (minus Vanderpool & Canto).

Studies included	K	N_h	Mean r	SD_r	%VE
All studies	23	8918	−.21	.354	1.9
Most-representative studies	14	4919	−.04	.347	2.4
Less-representative studies	9	3999	−.41	.231	2.9

Note. Bare-bones meta-analytical results: score differences between Amerindians and Whites, and *g* loadings. K = number of correlations; N = total sample size; mean r = mean-weighted vector correlation; SD_r = standard deviation of observed correlation; and %VE = percentage of variance accounted for by sampling errors.

mean value of $r(g \times d) = .62$ is a highly stable and reliable outcome, notwithstanding the fact that the majority of the variance between the data points still needs to be explained. Moreover, this value is virtually identical to the average value of .63 reported by Jensen (1998, p. 378) for a large number of studies carried out on Black/White comparisons using only high-quality samples.

Our results show that Spearman's hypothesis can be extended to the IQ gap between Amerindians and Whites — in other words, that the IQ gap is concentrated largely on *g*. Spearman's hypothesis applies to most cross-group comparisons thus examined and should be considered a regular phenomenon. However, the Amerindian IQ profile influences the outcomes: the Verbal subtests analyzed separately confirm Spearman's hypothesis, but the Performance subtests analyzed separately show a small anti-Jensen effect. Because of the massive *K*, this puzzling result is almost certainly not anomalous.

Our present study does, however, have limitations. We only used a limited sampling of tests in our examination of the Amerindian/White score gap: virtually all of our comparisons were on Wechsler tests. A more diverse sampling of tests is desirable when testing Spearman's hypothesis, whether pencil-and-paper/orally administered tests or chronometric batteries (Jensen, 1998, 2006).

In our paper we established a clear correlation between *g* loadings and group differences for Amerindians. The main contribution of this paper to the literature is in its empirical findings, and it is not necessary to discuss all aspects of the controversy surrounding Spearman's hypothesis. What makes a *g*-centered view of intelligence attractive for many researchers is that *g* is the latent variable that correlates best with most important real-world outcomes (see Jensen, 1998). However, see also van der Maas et al. (2006), Kan (2011), and Kan, Wicherts, Dolan, and van der Maas (2013) for a different interpretation of *g* loadings.

It is also possible that the *g* loadings of subtests correlate with the extent to which the subtest score correlates with educational attainment as measured by educational degrees or years of schooling, and that the direction of causality is from education to the subtests. For example, the WAIS-R subtest correlations with years in school reported by Kaufman, McLean, and Reynolds (1988) are similar to the *g* loadings of the subtests as reported, for example, in Table 2 of Kan et al. (2013). However, as tests of Spearman's hypothesis on individual datasets are of only limited strength, this hypothesis on educational loadings needs empirical testing on various datasets.

Jensen (1987) already discussed that subtest heritability can correlate with *g* loadings. A recent meta-analysis of Japanese studies shows a quite modest correlation, whereas a modest number of Western studies show a considerably higher correlation (te Nijenhuis, Kura, & Hur, 2014). It is clear that more studies are needed, which should be combined in a meta-analysis to establish the true value of the correlation between heritabilities and *g* loadings and where the influence of moderator variables could be empirically tested.

A reviewer came up with the suggestion that *g* loadings also correlate with subtest reliabilities. This is a hypothesis that can be tested empirically as well, preferably using data from various studies.

Acknowledgment

Thanks to Brian Stone for his help in locating studies. We would also like to thank Curt Hanson from the Department of Special Collections at the University of North Dakota for supplying us with papers from the Damian Vraniak archive.

References*

Armstrong, E. L., Woodley, M. A., & Lynn, R. (2014). Cognitive abilities amongst the Sámi people. *Intelligence*, *46*, 35–39.

Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The history and geography of human genes*. Princeton, NJ: Princeton University Press.

*Connelly, J. B. (1983). Recategorized WISC-R score patterns of older and younger referred Tlingit Indian children. *Psychology in the Schools*, *20*, 271–275.

*Cundick, B. P. (1970). Measures of intelligence on southwest Indian students. *The Journal of Social Psychology*, *81*, 151–156.

*Howell, R. J., Evans, L., & Downing, L. N. (1958). A comparison of test scores for the 16–17 year age group of Navaho Indians with standardized norms for the Wechsler Adult Intelligence Scale (Arizona and New Mexico). *The Journal of Social Psychology*, *47*, 355–359.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd Ed.). Thousand Oaks, CA: Sage.

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, *33*, 1–123.

Jensen, A. R. (1985). The nature of the black–white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences*, *8*, 193–263.

Jensen, A. R., & Faulstich, M. E. (1988). Difference between prisoners and the general population in psychometric g. *Personality and individual differences*, *9*, 925–928.

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

Jensen, A. R. (1987). The g beyond factor analysis. In R. R. Ronning, J. A. Glover, J. C. Conoley, & J. C. Witt (Eds.), *The influence of cognitive psychology on testing* (pp. 87–142). Hillsdale, NJ: Lawrence Erlbaum Associates.

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Amsterdam: The Netherlands: Elsevier.

*John, J. S., Krichev, A., & Bauman, E. (1976). Northwestern Ontario Indian children and the WISC. *Psychology in the Schools*, *13*, 407–411.

Kane, H. (2007). Group differences in nonverbal intelligence: Support for the influence of Spearman's g. *Mankind Quarterly*, *48*, 65–82.

Kan, K. -J. (2011). *The nature of nurture: The role of gene–environment interplay in the development of intelligence.* (PhD Dissertation) University of Amsterdam.

Kan, K. -J., Wicherts, J. M., Dolan, C. V., & van der Maas, H. L. J. (2013). On the nature and nurture of intelligence and specific cognitive abilities: The more heritable, the more culture-dependent. *Psychological Science*, *24*, 2420–2428.

Kaufman, A. S., McLean, J. E., & Reynolds, C. R. (1988). Sex, race, residence, region, and education differences on the 11 WAIS-R subtests. *Journal of Clinical Psychology*, *231*, 231–248.

Kush, J. C., & Watkins, M. W. (2007). Structural validity of the WISC-III for a national sample of Native American students. *Canadian Journal of School Psychology*, *22*, 235–248.

Lynn, R. (2006). *Race differences in intelligence: An evolutionary analysis.* Augusta, GA: Washington Summit Publishers.

* References marked with an asterisk were included into the meta-analysis.

Lynn, R. (2011). *The chosen people: A study of Jewish intelligence and achievement*. Washington, D.C.: Washington Summit.
McShane, D. A. (1980). Recategorized Wechsler scores of American Indian children. (Unpublished doctoral dissertation) George Peabody College for Teachers of Vanderbilt University, Tennessee.
*Mitchell, F. E. (1985). Assessing the performance differences of referred native American and non-native American children on two intelligence tests: K-ABC and WISC-R. (Unpublished doctoral dissertation) Oklahoma State University, Oklahoma.
*Nakano, S., & Watkins, M. W. (2013). Factor structure of the Wechsler Intelligence Scales for Children—Fourth edition among referred Native American students. *Psychology in the Schools*, *50*, 957–968.
Newman, J., Gallo, M. V., Schell, L. M., DeCaprio, A. P., Denham, M., & Deane, G. D. (2009). Analysis of PCB congeners related to cognitive functioning in adolescents. *Neurotoxicology*, *30*, 686–696.
*Reynolds, C. R., Willson, V. L., & Ramsey, M. (1999). Intellectual differences among Mexican Americans, Papagos and Whites, independent of g. *Personality and Individual Differences*, *27*, 1181–1187.
Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, *11*, 235–294.
Rushton, J. P., & Jensen, A. R. (2003). African–White IQ differences from Zimbabwe on the Wechsler Intelligence Scale for Children-Revised are mainly on the g factor. *Personality and Individual Differences*, *34*, 177–183.
Schmidt, F. L., & Le, H. (2004). *Software for the Hunter-Schmidt meta-analysis methods*. Department of Management & Organization, Iowa City, Iowa: University of Iowa 42242.
*Smith, S. D. (2009). A cross-cultural analysis of tribal differences of American Indian children on the Wechsler Intelligence Scale for Children-III (WISC-III). (Unpublished doctoral dissertation) Capella University, Minnesota.
*Snyder, B. J. (1991). WISC-R Performance patterns of referred Anglo, Hispanic, and American Indian Children. (Unpublished doctoral dissertation). University of Arizona, Arizona.
Statistics Canada (2011). *National Household Survey*. Statistics Canada Catalogue no. 99–010-X2011028. (Retrieved on November, 18, 2014, from http://www12.statcan.gc.ca/nhs-enm/2011/).
Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.
*Teeter, A., Moore, C. L., & Petersen, J. D. (1982). WISC-R verbal and performance abilities of Native American students referred for school learning problems. *Psychology in the Schools*, *19*, 39–44.
te Nijenhuis, J., & van der Flier, H. (1997). Comparability of GATB scores for immigrants and majority group members: Some Dutch findings. *Journal of Applied Psychology*, *82*, 675–687.
te Nijenhuis, J., & van der Flier, H. (2003). Immigrant–majority group differences in cognitive performance: Jensen effects, cultural effects, or both? *Intelligence*, *31*, 443–459.
te Nijenhuis, J., Kura, K., & Hur, Y. M. (2014). The correlation between g loadings and heritability in Japan: A meta-analysis. *Intelligence*, *46*, 275–282.
*Turner, G. H., & Penfold, D. J. (1952). The scholastic aptitude of Indian children of the Caradoc Reserve. *Canadian Journal of Psychology*, *6*, 31–44.
U.S. Census Bureau (2007). *The American Community—American Indians and Alaska Natives: 2004*. Washington, DC: US Census Bureau. (Retrieved November 10, 2014, from http://www.census.gov/prod/2007pubs/acs-07.pdf.).
U.S. Census Bureau (2012). *The American Indian and Alaska Native Population: 2010*. Washington, DC: US Census Bureau. (Retrieved November 14, 2014, from http://www.census.gov/prod/cen2010/briefs/c2010br-10.pdf.).
van der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842.
*Vanderpool, M., & Catano, V. M. (2008). Comparing the performance of Native North Americans and predominantly White military recruits on verbal and nonverbal measures of cognitive ability. *International Journal of Selection and Assessment*, *16*, 239–248.