Spearman's Law of Diminishing Returns in Normative Samples for the WISC-IV and WAIS-III

Donald H. Saklofske[1], Zhiming Yang[2], Jianjun Zhu[2], and Elizabeth J. Austin[3]

[1]University of Calgary, Canada, [2]Psychological Corporation, San Antonio, TX, USA, [3]University of Edinburgh, Scotland, UK

Abstract. In order to explain observed variations in intelligence test scores, Spearman (1927) proposed the "law of diminishing returns" (SLODR). It states that the *g* saturation of cognitive ability tests decreases as a function of ability or age. Published studies have shown mixed results. However, a recent review (Hartmann & Nyborg, 2004) suggests that there is evidence for differences in *g* saturation by ability level, but that observed age effects on *g* saturation are most likely to be a consequence of the ability effect. The current study analyzed the standardization data of the most recent Wechsler scales for both children and adults from several different countries. This study did not find evidence to support either the ability or age version of SLODR by using large normative samples for the WISC-IV from the United States, Canada, and Australia, and for the WAIS-III from the same three countries and also from The Netherlands.

Keywords: intelligence, psychometrics, Spearman's law, Wechsler tests

In 1927, Spearman developed a two-factor model to explain the variation in intelligence test scores. The specific factor (*s*) makes a person more skilled at one cognitive task than another while a general factor (*g*) governs performance on all cognitive tasks. Spearman's "law of diminishing returns" (SLODR; Spearman, 1927) states that the *g* saturation of cognitive ability tests decreases as a function of ability or age. Thus, the *g* loadings of test scores are expected to be higher for low ability or young groups, with larger test/subtest correlations being found in these groups. This effect is also referred to as *ability differentiation* as a function of age and ability level.

Focusing on ability level effects, one theoretical approach to SLODR is to employ a biologically-based argument that, at lower levels of *g*, central-processor efficiency limits the possibility of differentiation for the less able. Thus, they would be expected to perform uniformly poorly on all types of ability tests. Anderson (1992) proposed that, at lower ability levels, performance is essentially determined by basic processing speed, whereas at higher ability levels/processing speed, differences in specific processors related to particular tasks should become more apparent. An alternative perspective on SLODR is that of the investment model of intelligence (Cattell, 1987), stating that, at higher levels of *g*, individuals are more free to follow developmental trajectories that result in cognitive resources being invested in diverse ways in specific factors, for example by combining excellent mathematical skills with more modest verbal skills or vice versa.

The idea of *g*-dependent developmental trajectories links to a genetically-based model of SLODR, which proposes that the heritability of *g* is lower at higher ability levels, thus, allowing for a larger contribution from environmental effects, with larger nonshared environmental effects leading to larger differences in developmental trajectories. Studies to date on the variation of the heritability of *g* with ability have, however, produced contradictory results (Bailey & Revelle, 1991; Cherny, Cardon, Fulker, & DeFries, 1992; Detterman, Thomson, & Plomin, 1990; Jensen, 1997), leaving the status of this genetic model unclear.

The theory that mental traits act as sexually selected fitness indicators (Miller, 2000), a view supported by findings of associations between ability test scores and body symmetry (Bates, 2007; Furlow, Armijo-Prewitt, Gangestad, & Thornhill, 1997; Prokosch, Yeo, & Miller, 2005), also has possible links to SLODR. Miller (2000) suggested a process analogous to the investment model in which different individuals specialize in different preferred types of fitness display (mental ability), with the implication that this could be more marked in the more able.

An alternative, psychometric perspective on SLODR is that test characteristics may be relevant, with ceiling effects (because the tests are easy and, hence, poorly discriminating for the most able) reducing test-test correlations for the high-ability compared to low-ability groups (Fogarty & Stankov, 1995). From this viewpoint SLODR is simply an artifact of the use of tests designed to be optimally discriminating at ability levels close to the population mean. Using more difficult or cognitively challenging tests optimized for the highly able should reverse the effect. It has also been suggested that

DOI 10.1027/1614-0001.29.2.57

cognitive ability differentiation may be an outcome of education (Abad, Colom, Juan-Espinosa, & Garcia, 2003; Colom, Abad, Garcia, & Juan-Espinosa, 2002).

As regards ability differences in *g*-loadings, a number of recent studies have confirmed the effect (e.g., Abad et al., 2003; Deary et al, 1996; Detterman & Daniel, 1989; Legree, Pifer, & Grafton, 1996; Jensen, 2003; te Nijenhuis & Hartmann, 2006), including a study of the strength of association between reaction times and test scores (Der & Deary, 2003), although there are also studies that have not confirmed this (e.g., Facon, 2004; Fogarty & Stankov, 1995; Hartmann & Reuter, 2006; Hartmann & Teasdale, 2004, 2005). Contradictory findings have been attributed to the small effect size found for SLODR; if this effect is weak many studies would lack sufficient statistical power to detect the associated small changes in *g*-loadings (Hartmann & Nyborg, 2004; te Nijenhuis & Hartmann, 2006). Further, some null findings can be accounted for by the requirement that for the observation of SLODR a test battery should be sufficiently diverse (both in terms of the number of tests and in terms of including tests covering a wide range of ability subcomponents) for specific as well as general factors to emerge from it. Batteries containing only a small number of tests and/or containing only tests of similar content are unlikely to exhibit SLODR (Hartmann & Teasdale, 2005). The review by Hartmann and Nyborg (2004) of virtually all published studies on SLODR, including a careful examination of the validity of each study, concludes that there is support for small differences in *g* saturation across ability groups.

The contradictory findings about SLODR appear to be related to several research design issues, such as the test battery selection, the sample splitting method, the score range for each ability group, the statistical methods, and so on. For example, using age-corrected standard scores or raw total scores in the analysis may lead to different conclusions (Detterman & Daniel, 1989; Jensen, 2003; Hartmann & Teasdale, 2004, 2005; Hartmann & Reuter, 2006).

For age effects where comparisons are made across childhood/early adult years, there have been recent mixed findings of confirmation (Tideman & Gustafsson, 2004) and, more often, nonconfirmation of SLODR (Carroll, 1993; Facon, 2004; Hartmann, 2006; Juan-Espinosa, García, Colom, & Abad, 2000; Rietveld, Dolan, van Baal, & Boomsma, 2003), although earlier studies were more supportive of the effect (e.g., Bayley, 1955; Burt, 1954; Garrett, 1946). Interestingly, recent findings are somewhat more supportive of a related effect, proposed by cognitive aging researchers, that abilities "de-differentiate" in old age because of a decline in processing efficiency (e.g., Ghisletta & Lindenberger, 2003). Hartmann and Nyborg (2004) concluded that change in *g* saturation with age "follows the growth curve for general ability and is, furthermore, eliminated when controlling for ability. This points to the fact that SLODR (age) is nothing more than a by-product of differences in ability and consequently SLODR (ability)" (p. 114).

The purpose of this study was to further investigate SLODR. The normative samples across the United States, Canada, and Australia from the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; Wechsler, 2003), and the normative samples across the United States, Canada, Australia, and the Netherlands from the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III; Wechsler, 1997) were analyzed by exploratory factor analysis (EFA). The Wechsler scales are established and often-used measures of cognitive ability (Tulsky, et al., 2003) and yield a general mental ability score derived from a number of subtests (Kaufman, Flanagan, Alfonso, & Mascolo, 2006). These more recent versions of both the adult and children's scales have been standardized in a number of English-speaking countries as well as translated and normed for use in non-English-speaking countries (e.g., see Georgas, Weiss, van de Vijver, & Saklofske, 2003). For this study, both classical test theory (CTT) and item response theory (IRT) were used to calibrate or estimate ability scores. These IRT-based and CTT-based absolute ability scores were used in sample splitting and factor analysis in this study.

Study 1: *g* Loadings on the WISC-IV

Method

Instrument

The WISC-IV contains 15 core and supplementary subtests that assess verbal comprehension, perceptual reasoning, working memory, and processing speed as well as full-scale IQ. Only 8 of the 10 core subtests were used in this study and include similarities (SI), vocabulary (VC), comprehension (CO), block design (BD), picture concepts (PS), matrix reasoning (MR), digit span (DS), and letter-number sequencing (LNS). Two of the core subtests, coding (CD) and symbol search (SS), were excluded from this analysis because they are processing speed tests and no IRT-based ability scores can be estimated for them.

Participants

The WISC-IV normative samples of children, ages 6–16 years, from the United States (N = 2200), Canada (N = 1100), and Australia (N = 851) were used in this study.

Statistical Analyses

Since confirmatory factor analysis (CFA) and EFA yielded identical *g*-loading patterns using the WISC-IV and the WAIS-III US normative sample data, EFA was used to estimate Spearman's general factor (g) in this study. Usually, the first unrotated factor in the principal axis EFA can be considered as a general factor because most subtests (variables) load highly on it (Sattler, 2001, p. 119). A simulation study (Jensen & Weng, 1994) found that this method was one (of several) which gives a good estimate of *g*.

To investigate the occurrence of any changes in *g* loadings across age or ability groups, the WISC-IV normative sample was split into two groups following the classical SLODR analysis method (Hartmann & Teasdale, 2004) whereby two groups were constructed by applying an equal-variance criterion to determine the cut-off point on the average IRT-based ability score or the convenient age band. The IRT ability score is a derived equal interval score with a mean of 500 and *SD* of 10 and is determined using a Rasch model concurrent calibration method. Comparing to the norm-referenced standard scores, using IRT ability scores to split the sample into higher/lower ability groups is appropriate because the IRT ability scores are located on a unique scale across all age or ability groups. Actually, the IRT ability score is a kind of absolute ability score: The higher/lower ability groups based on IRT ability scores are absolute higher/lower ability groups. In contrast, since the standard scores are age-corrected norm-referenced scores, the higher/lower ability groups based on age-corrected standard scores are not absolute higher/lower ability groups. For example, the ability level of an IQ 82, 15-year-old child may be much higher than the ability level of an IQ 108, 7-year-old child. Putting different age group's lower IQ children together does not allow the construction of an absolute lower ability group. Compared to the subtest raw scores, IRT ability scores are better than raw scores for splitting the sample into higher/lower ability groups because the IRT ability scores are equal interval scores. More ability groups or age groups were not used in this study because of the restriction of score-range problem (Hunter & Schmidt, 2004; Jensen, 2003) and the resulting small sample-size problem for the extreme ability or age groups while ensuring equal variances across multiple groups.

In order to estimate the overall effect of *g*-loading changes between two groups for the whole test battery, the eigenvalues were transformed into average intercorrelations using Kaiser's formula (Kaiser, 1968):

$$\gamma = \frac{\lambda - 1}{p - 1}.$$

Here, λ is the first unrotated principal component (PC1), p is the number of included subtests in the factor analysis, and γ is the average intercorrelation. The actual effect sizes were calculated by subtracting the average intercorrelation of the higher ability/older age group from the average intercorrelation of the lower ability/young age group. If the γ difference is negative, then the SLODR (age/ability) is confirmed. Variance ratios (*F*) of the PC1 eigenvalues for Lower/Higher ability or Young/Old age group were estimated. Since the *n* count is relatively large, the Cohen's *d* effect sizes (Hays, 1994, p. 411) were also calculated for testing the *g*-loading changes.

$$d = \sqrt{F\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}.$$

Here N_1 and N_2 are the sample sizes for two groups in the *F* test. Cohen (1988) also defined effect sizes as small, $d = .2$; medium, $d = .5$; and large, $d = .8$.

In addition to examining the average intercorrelation of all subtests, the individual subtest *g*-loading changes were also investigated in this study.

Results

g Loadings Across Age Groups

In order to investigate the changes of *g* loading across age groups, the WISC-IV normative sample was divided into the young age group (age 6–8 years) and the older age group (age 9–16 years). Correlations between age and ability for these groups and for the whole sample are shown in Table 1.1 It can be seen that age and ability scores are highly correlated, ranging from 0.56 to .78 for the overall normative samples across the three countries.

The IRT ability score mean and *SD* between the young and the older age groups for each subtest, based on data from the three countries, are presented in Table 1.2. The *F* testing for the average variances indicated that there are no significant differences between the two group variances ($p > .10$, two-tailed). Please note that the *F* ratios for all subtests in this table can be ignored because any small difference between group variances will be identified as a significant difference because of the large sample size. The effect size (Cohen's *d*), which was calculated using Hays' formula (Hays, 1994, p. 411), was used to evaluate the variance difference because *d* values are free from sample size effects. From Table 1.2, the values of effect size *d* for subtests and the average are very small, ranging from 0.0485 to 0.0980. These results demonstrated that this study met the equal variance assumption.

The WISC-IV *g* loadings across age groups for the United States, Canada, and Australia normative sample are presented in Table 1.3.

From Table 1.3, the *g* loadings of the SI, VC, CO, BD, and LNS subtests are higher for the 9–16-year-old age group of children across the United States, Canada, and Australia samples. The *g* loadings of the MR and DS subtests are also higher for older children in both the US and Australian samples but not in the Canadian sample. The *g* loading of the PS subtest, however, is higher for the 6–8-year-old age group in both the US and Australian samples. Table 1.3 also indicates that the average intercorrelation (γ), which is a *g*-loading effect size for the whole test battery, is increased from the young age groups to the older age groups. However, the Cohen's *d* effect sizes (Hays, 1994, p. 411) are small, ranging from .0497 to .0797. These findings do not provide evidence to support Spearman's hy-

Table 1.1. Correlations between age and ability for WISC-IV normative sample, overall, and by age group

Subtest	USA Overall	Age6-8	Age9-16	Canada Overall	Age6-8	Age9-16	Australia Overall	Age6-8	Age9-16
SI	0.68	0.44	0.44	0.70	0.35	0.48	0.70	0.39	0.47
VC	0.73	0.52	0.52	0.74	0.40	0.54	0.78	0.56	0.57
CO	0.73	0.47	0.53	0.76	0.41	0.57	0.73	0.45	0.51
BD	0.66	0.33	0.46	0.67	0.37	0.45	0.71	0.44	0.54
PS	0.60	0.47	0.33	0.62	0.41	0.36	0.57	0.47	0.29
MR	0.64	0.49	0.34	0.65	0.44	0.35	0.64	0.48	0.40
DS	0.56	0.37	0.33	0.59	0.35	0.36	0.57	0.37	0.38
LN	0.63	0.42	0.39	0.66	0.41	0.42	0.65	0.50	0.42
N	2200	600	1600	1100	300	800	851	235	616

Note: SI = Similarity, VC = Vocabulary, CO = Comprehension, BD = Block design, PS = Picture concepts, MR = Matrix reasoning, DS = Digit span, LN = Letter-number sequence.

Table 1.2. The IRT-ability score mean and *SD* across ability groups with WISC-IV normative samples

Sub-test[1]	USA Age 6–8 Mean	*SD*	Age 9–16 Mean	*SD*	*F*[2]	*d*[3]	Canada Age 6–8 Mean	*SD*	Age 9–16 Mean	*SD*	*F*	*d*	Australia Age 6–8 Mean	*SD*	Age 9–16 Mean	*SD*	*F*	*d*
SI	490.03	8.48	503.69	7.68	1.22	0.0528	489.82	8.57	503.87	7.44	1.33	0.0551	489.35	9.07	503.94	7.10	1.63	0.0980
VC	489.92	6.93	503.85	8.25	1.42	0.0570	490.18	7.14	503.77	7.05	1.03	0.0485	488.86	6.45	504.13	7.67	1.42	0.0913
CO	489.92	6.74	503.74	8.28	1.51	0.0588	489.29	7.12	504.01	7.68	1.16	0.0517	489.44	7.69	504.01	7.45	1.07	0.0791
BD	490.66	7.79	503.46	8.42	1.17	0.0518	490.58	7.66	503.63	8.23	1.15	0.0514	490.58	7.41	503.62	8.32	1.26	0.0862
PS	490.74	9.14	503.41	7.97	1.32	0.0549	490.54	9.06	503.47	7.80	1.35	0.0556	491.19	9.67	503.26	7.95	1.48	0.0933
MR	489.87	8.97	503.83	7.43	1.46	0.0578	489.61	8.92	503.82	7.16	1.55	0.0596	490.68	9.30	503.64	7.71	1.46	0.0925
DS	491.89	8.15	502.99	8.81	1.17	0.0517	491.41	7.92	503.11	8.84	1.25	0.0535	492.26	8.30	502.95	9.00	1.17	0.0831
LN	490.96	7.77	503.35	8.57	1.22	0.0528	490.34	9.04	503.67	7.60	1.41	0.0569	490.78	8.00	503.42	8.37	1.09	0.0802
Average[4]	490.50	6.10	503.54	6.43	1.04	0.0488	490.22	6.26	503.67	6.06	1.12	0.0508	490.39	5.95	503.62	5.87	1.09	0.0799
N	600		1600				300		800				235		616			

Note: [1]SI = Similarity, VC = Vocabulary, CO = Comprehension, BD = Block design, PS = Picture Concepts, MR = Matrix reasoning, DS = Digit span, LN = Letter-Number Sequence. [2]*F* testing for each subtest may be misleading because very little difference will be identified as significant difference due to big sample size. [3]*d* is effect size. [4]*F* variance ratio for the mean of the eight subtest variances is calculated using the following method (Hartmann & Teasdale, 2005, p. 1197): $SD^2_{Average(Larger)}/SD^2_{Average(Smaller)}$, where $SD^2_{Average} = (SD^2_{SI} + SD^2_{VC} + SD^2_{CO} + SD^2_{BD} + SD^2_{PS} + SD^2_{MR} + SD^2_{DS} + SD^2_{LN})$ and *F* ratio for the average is 1.04, 1.12, and 1.09 for US, Canadian, and Australian samples, respectively ($p > .10$, two tailed).

Table 1.3. The *g* loadings across age groups with WISC-IV normative samples

Subtest*	USA Age 6–8	Age 9–16		Canada Age 6–8	Age 9–16		Australia Age 6–8	Age 9–16	
SI	0.8255	0.8551	+	0.8230	0.8605	+	0.7135	0.8144	+
VC	0.8476	0.8886	+	0.8485	0.8700	+	0.8091	0.8600	+
CO	0.7604	0.8183	+	0.7673	0.8381	+	0.6306	0.7638	+
BD	0.6438	0.7323	+	0.6579	0.6815	+	0.6666	0.7053	+
PS	0.6852	0.6566	–	0.6620	0.6760	+	0.6949	0.5379	–
MR	0.7408	0.7500	+	0.7025	0.6835	–	0.6555	0.6620	+
DS	0.6355	0.6486	+	0.6789	0.6673	–	0.6064	0.6104	+
LN	0.6753	0.7076	+	0.7048	0.7704	+	0.6580	0.6621	+
Eigenvalue of the PC1	4.27	4.64	$F = 1.08$	4.31	4.62	$F = 1.07$	3.72	4.02	$F = 1.08$
Average intercorrelation (γ)	.47	.52	$d = .0497$	.47	.52	$d = .0700$	.39	.43	$d = .0797$
N	600	1600		300	800		235	616	

Note: * SI = Similarity, VC = Vocabulary, CO = Comprehension, BD = Block design, PS = Picture Concepts, MR = Matrix reasoning, DS = Digit span, LN = Letter-number sequence. +: *g* loadings increased from young to old group. –: *g* loadings decreased from young to old group.

Table 2.1. Correlations between age and ability for WISC-IV normative sample, overall, and by ability group

	USA			Canada			Australia		
Subtest	Overall	Low	High	Overall	Low	High	Overall	Low	High
SI	0.68	0.20	0.49	0.70	0.17	0.55	0.70	0.32	0.51
VC	0.73	0.23	0.57	0.74	0.25	0.61	0.78	0.50	0.61
CO	0.73	0.24	0.60	0.76	0.22	0.65	0.73	0.42	0.55
BD	0.66	0.17	0.50	0.67	0.20	0.51	0.71	0.30	0.54
PS	0.60	0.25	0.34	0.62	0.16	0.39	0.57	0.29	0.25
MR	0.64	0.22	0.38	0.65	0.23	0.43	0.64	0.27	0.38
DS	0.56	0.05	0.34	0.59	0.14	0.37	0.57	0.10	0.36
LN	0.63	0.19	0.39	0.66	0.09	0.46	0.65	0.30	0.41
N	2200	577	1623	1100	283	817	851	233	618

Note: SI = Similarity, VC = Vocabulary, CO = Comprehension, BD = Block design, PS = Picture concepts, MR = Matrix reasoning, DS = Digit span, LN = Letter-number sequence.

Table 2.2. The IRT-ability score mean and *SD* across ability groups with WISC-IV normative samples

	USA						Canada						Australia					
Subtest[1]	Low		High				Low		High				Low		High			
	Mean	*SD*	Mean	*SD*	*F*[2]	*d*[3]	Mean	*SD*	Mean	*SD*	*F*	*d*	Mean	*SD*	Mean	*SD*	F	*d*
SI	487.89	7.87	504.25	6.53	1.45	0.0583	488.12	8.15	504.17	6.68	1.49	0.0841	488.12	8.82	504.36	6.22	2.01	0.1090
VC	488.06	6.43	504.31	7.28	1.28	0.0548	488.90	6.67	503.93	6.57	1.03	0.0701	488.15	5.92	504.35	7.32	1.53	0.0952
CO	488.54	6.77	504.04	7.52	1.23	0.0539	488.07	6.91	504.13	7.17	1.07	0.0715	488.79	7.34	504.21	7.14	1.06	0.0790
BD	489.15	7.74	503.82	7.66	1.02	0.0490	489.52	7.69	503.72	7.80	1.03	0.0700	489.24	6.98	504.08	7.62	1.19	0.0840
PS	488.91	8.50	503.88	7.22	1.39	0.0571	488.57	8.33	503.88	7.07	1.39	0.0812	489.33	8.34	503.92	7.35	1.29	0.0872
MR	487.51	8.16	504.47	6.13	1.77	0.0645	488.06	8.40	504.06	6.56	1.64	0.0884	488.60	8.27	504.38	6.65	1.55	0.0957
DS	489.76	7.77	503.60	7.90	1.03	0.0493	489.54	7.46	503.52	8.16	1.20	0.0755	490.61	7.81	503.54	8.34	1.14	0.0820
LN	488.90	6.82	503.91	7.76	1.29	0.0552	487.83	8.11	504.26	6.47	1.57	0.0865	489.24	7.29	503.96	7.67	1.11	0.0809
Average[4]	488.59	5.10	504.03	5.33	1.07	0.0502	488.58	5.47	503.96	5.24	1.19	0.0754	489.01	4.77	504.10	5.00	1.09	0.0803
N	577		1623				283		817				233		618			

Note: [1]SI = Similarity, VC = Vocabulary, CO = Comprehension, BD = Block design, PS = Picture concepts, MR = Matrix reasoning, DS = Digit span, LN = Letter-number sequence. [2]*F* testing for each subtest may be misleading because very little differences are identified as significant differences because of the large sample size. [3]*d* = effect size. [4]*F* variance ratio for the mean of the eight subtest variances is calculated using the following method (Hartmann & Teasdale, 2005, p. 1197):
$SD^2_{Average(Larger)}/SD^2_{Average(Smaller)}$, where $SD^2_{Average} = (SD^2_{SI} + SD^2_{VC} + SD^2_{CO} + SD^2_{BD} + SD^2_{PS} + SD^2_{MR} + SD^2_{DS} + SD^2_{LN})$ and *F* ratio for the average is 1.07, 1.19, and 1.09 for USA, Canada, and Australia sample, respectively ($p > .10$, two-tailed).

Table 2.3. The *g* loadings across ability groups with WISC-IV normative samples

	USA			Canada			Australia		
Subtest*	Low	High		Low	High		Low	High	
SI	0.7480	0.8168	+	0.7807	0.8343	+	0.6340	0.8044	+
VC	0.7954	0.8683	+	0.8111	0.8602	+	0.7841	0.8521	+
CO	0.7152	0.7915	+	0.7279	0.8192	+	0.5462	0.7550	+
BD	0.5580	0.6865	+	0.6241	0.6482	+	0.5170	0.6545	+
PS	0.5920	0.5655	–	0.5843	0.5812	–	0.5541	0.4075	–
MR	0.6091	0.6536	+	0.6271	0.6122	–	0.4858	0.5396	+
DS	0.5437	0.5412	–	0.5920	0.5987	+	0.4690	0.5255	+
LN	0.4537	0.6536	+	0.5711	0.6876	+	0.5400	0.5759	+
Eigenvalue of the PC1	3.24	3.99	*F* = 1.23	3.60	4.07	*F* = 1.13	2.64	3.44	*F* = 1.30
Average intercorrelation(γ)	.32	.43	*d* = .0538	.37	.43	*d* = .0733	.23	.35	*d* = .0877
N	577	1623		283	817		233	618	

Note: *SI = Similarity, VC = Vocabulary, CO = Comprehension, BD = Block design, PS = Picture Concepts, MR = Matrix reasoning, DS = Digit span, LN = Letter-Number Sequence. +: *g* loadings increased from lower to higher ability group. –: *g* loadings decreased from lower to higher ability group.

pothesis that the *g* saturation of cognitive ability tests decreases as a function of age.

g Loadings Across Ability Groups

In the analysis of the *g* loadings across ability groups, the WISC-IV normative sample was divided into a low ability group and a high ability group according to IRT-based ability scores. The cut-off point was 495 in this study. Correlations between age and ability for ability groups and for the whole sample are shown in Table 2.1. It can be seen that age and ability scores are highly correlated for the overall normative samples. Compared to the high ability group, the correlations between age and ability for the low ability group are relatively lower because of the restriction of ability score range.

The IRT ability score mean and *SD* between the lower and the higher ability groups for each subtest are presented in Table 2.2. The *F* testing for the average variances indicated that there are no significant differences between the two ability group variances ($p > .10$). For each subtest, the *F* ratio can also be ignored because any small variance difference between groups will be identified as a significant difference because of the large sample size. The Cohen's *d* effect sizes (Hays, 1994, p. 411) for testing score variance differences between the two ability groups for both the average and subtests are small, ranging from 0.0490 to 0.1090. These results showed that the assumption of equal ability-score variances between ability groups is met.

The WISC-IV *g* loadings across ability groups drawn from the United States, Canada, and Australia normative samples are presented in Table 2.3.

The *g* loadings of SI, VC, CO, BD, and LNS are not lower for the higher ability group in any of the US, Canadian, and Australian samples. The *g* loading of the PS subtest is higher for the lower ability group across the US, Canadian, and Australian samples. For the whole test battery, Table 2.3 also demonstrates that the average intercorrelation (γ) is increased from the lower ability group to the higher ability group. Again, the Cohen's *d* effect sizes (Hays, 1994, p. 411) are small, ranging from .0538 to .0877. This result does not support Spearman's hypothesis that the *g* saturation of cognitive ability tests decreases as a function of ability.

Study 2: *g* Loadings on the WAIS-III

Method

Instrument

The WAIS-III is used with examinees aged 16–89 years. The 13 subtests assess the broad factors of Verbal comprehension, Perceptual organization, Working memory, and Processing speed, as well as providing a full scale IQ. Only 12 subtests were used in this study and included vocabulary (VOC), similarities (SIM), information (INF), comprehension (COM), picture completion (PIC), block design (BLD), matrix reasoning (MR), picture arrangement (PA), arithmetic (ARI), digit span (DS), digit symbol-coding (DAS), and symbol search (SYS). The letter-number sequencing subtest was excluded because of missing values in the WAIS-III US normative sample ($N = 1{,}250$ instead of 2,450).

Participants

The WAIS-III normative samples from the United States ($N = 2{,}450$), Canada ($N = 1{,}105$), Australia ($N = 297$), and Netherlands ($N = 670$) were used in this study.

Statistical Analyses

The principal axis EFA of the WAIS-III was conducted with CTT ability scores. The CTT ability score used here was a derived score with a mean of 500 and *SD* of 10 from the corresponding *z* score based on the overall normative sample. In contrast to the WISC-IV study reported above, the IRT ability score was not used because the item level scores were not available for all of the WAIS-III data sets used here. The first unrotated factor loadings were treated as the general factor loadings (Sattler, 2001, p. 119). In order to investigate the changes of *g* loadings across age or ability groups, the WAIS-III normative sample was divided into two groups based on CTT ability scores or age by controlling for equal variances between the groups.

Results

g Loadings Across Age Groups

In the age-group study, the WAIS-III normative sample was split into the young (16–44 years) and the older age group (45 years and older). Correlations between age and ability for these groups and for the whole sample are shown in Table 3.1. For the WAIS-III US normative sample, the correlations for half of the subtests are above 0.42 for the overall sample. Compared to the older age group, the correlations for the young age group are relatively small. This indicates that the cognitive abilities of young adults are more stable than those of older adults. The negative correlations for the older age group also indicate that ability declines after age 45.

The CTT ability score mean and *SD* between the young and the older age groups are presented in Table 3.2. The *F* testing for the average variance indicated that there were no significant differences between group variances ($p > .10$). For each subtest, the *F* ratio can be ignored because any small difference between group variances will be identified as significant because of the large sample size. The Cohen's d effect sizes (Hays, 1994, p. 411) of the ability-score variance differences between the two age groups for both the average and subtests are reported in Table 3.2. The d values are small and range from 0.0405 to 0.1496. These results indicated that the assumption of equal variances between age groups was met in this study.

The WAIS-III g loadings across age groups for the US, Canada, Australia, and Netherlands normative samples are presented in Table 3.3.

Table 3.1. Correlations between age and ability for WAIS-III normative sample, overall, and by age group

	USA			Canada			Australia			Netherlands		
Subtest	Overall	Age16-44	Age45-89	Overall	Age16-44	Age45-80	Overall	Age16-44	Age45-89	Overall	Age16-44	Age45-86
VOC	0.06	0.26	–0.17	–0.03	0.27	–0.21	0.05	0.28	–0.14	0.17	0.27	0.03
SIM	–0.21	0.15	–0.29	–0.28	0.13	–0.41	–0.27	0.08	–0.30	–0.10	0.09	–0.16
INF	0.05	0.13	–0.21	–0.01	0.19	–0.24	0.04	0.21	–0.12	0.06	0.10	–0.02
COM	0.02	0.22	–0.24	–0.09	0.24	–0.30	0.05	0.28	–0.15	0.09	0.17	–0.09
PIC	–0.42	–0.01	–0.37	–0.38	0.05	–0.49	–0.35	0.13	–0.40	–0.37	–0.02	–0.40
BLD	–0.48	–0.01	–0.40	–0.49	–0.06	–0.51	–0.54	–0.18	–0.46	–0.43	–0.08	–0.40
MR	–0.55	–0.13	–0.41	–0.51	–0.09	–0.54	–0.51	–0.17	–0.44	–0.47	–0.10	–0.35
PA	–0.57	–0.07	–0.48	–0.55	–0.09	–0.53	–0.49	0.02	–0.52	–0.56	–0.10	–0.37
ARI	–0.13	0.10	–0.27	–0.18	0.10	–0.38	–0.15	0.01	–0.19	–0.04	0.08	0.04
DS	–0.25	–0.04	–0.22	–0.25	–0.05	–0.27	–0.25	–0.11	–0.19	–0.18	–0.08	0.03
DAS	–0.65	–0.08	–0.56	–0.54	–0.15	–0.54	–0.30	0.01	–0.37	–0.56	–0.04	–0.35
SYS	–0.62	–0.10	–0.53	–0.62	–0.11	–0.61	–0.55	–0.17	–0.48	–0.57	–0.10	–0.36
N	2450	1195	1255	1105	652	453	297	170	127	670	338	332

Note: VOC = Vocabulary, SIM = Similarity, INF = Information, COM = Comprehension, PIC = Picture completion, BLD = Block design, MR = Matrix Reason, PA = Picture arrangement, ARI = Arithmetic, DS = Digit span, DAS = Digit symbol-coding, SYS = Symbol search.

Table 3.2. The CTT-Ability score mean and *SD* across age groups with WAIS-III normative samples

	USA						Canada					
	Age16-44		Age45-89		F[2]	d[3]	Age16-44		Age45-80		F	d
Subtest[1]	Mean	*SD*	Mean	*SD*			Mean	*SD*	Mean	*SD*		
VOC	499.23	9.81	500.69	10.12	1.06	0.0417	500.29	9.49	499.54	10.79	1.29	0.0696
SIM	501.76	9.44	498.35	10.36	1.20	0.0443	501.92	8.84	497.17	10.73	1.47	0.0743
INF	499.09	9.93	500.89	9.91	1.00	0.0405	500.03	9.83	500.05	10.34	1.11	0.0643
COM	499.39	9.94	500.52	9.97	1.01	0.0405	500.64	9.58	499.13	10.49	1.20	0.0669
PIC	503.46	7.83	496.70	10.73	1.88	0.0554	502.51	7.91	496.54	11.53	2.12	0.0892
BLD	504.37	9.64	495.92	8.38	1.32	0.0465	503.39	9.05	495.06	9.21	1.04	0.0622
MR	504.92	8.56	495.41	8.94	1.09	0.0422	503.40	8.20	495.17	10.31	1.58	0.0770
PA	504.94	7.97	495.31	9.40	1.39	0.0477	503.81	7.68	494.38	10.19	1.76	0.0812
ARI	500.85	10.11	499.19	9.92	1.04	0.0412	501.08	9.81	498.54	10.15	1.07	0.0633
DS	502.20	10.08	498.01	9.52	1.12	0.0428	501.56	9.79	497.52	10.09	1.06	0.0630
DAS	505.76	8.04	494.53	8.54	1.13	0.0430	503.64	9.66	493.44	10.54	1.19	0.0668
SYS	505.63	8.56	494.69	8.25	1.08	0.0419	504.62	8.07	493.73	8.70	1.16	0.0659
Average[4]	502.63	6.83	497.52	7.41	1.07	0.0419	502.24	6.06	496.69	7.59	1.30	0.0697
N	1195		1255				652		453			

Note: [1]VOC = Vocabulary, SIM = Similarity, INF = Information, COM = Comprehension, PIC = Picture completion, BLD = Block design, MR = Matrix reason, PA = Picture arrangement, ARI = Arithmetic, DS = Digit span, DAS = Digit symbol-coding, SYS = Symbol search. [2]F testing for each subtest may be misleading because very little difference will be identified as significant difference due to big sample size. [3]d is effect size. [4]F variance ratio for the mean of the eight subtest variances is calculated using the following method (Hartmann & Teasdale, 2005, p. 1197):

$F = SD^2_{Average(Larger)} / SD^2_{Average(Smaller)}$, where

$SD^2_{Average} = (SD^2_{VOC} + SD^2_{SIM} + SD^2_{INF} + SD^2_{COM} + SD^2_{PIC} + SD^2_{BLD} + SD^2_{MR} + SD^2_{PA} + SD^2_{ARI} + SD^2_{DS} + SD^2_{DAS} + SD^{2)}_{SYS}$ and F ratio for the average is 1.07, 1.30, 1.11, and 1.26 for USA, Canada, Australia and Netherlands sample, respectively ($p > .10$, two-tailed).

Table 3.2. (continued)

Subtest	Australia Age16-44 Mean	SD	Age45-89 Mean	SD	F	d	Netherlands Age16-44 Mean	SD	Age45-86 Mean	SD	F	d
VOC	499.82	10.05	500.30	9.91	1.03	0.1190	498.80	9.94	501.30	9.95	1.00	0.0773
SIM	502.31	8.78	497.49	10.68	1.48	0.1426	500.83	10.24	499.02	10.08	1.03	0.0785
INF	499.76	10.00	500.32	10.09	1.02	0.1183	499.50	10.09	500.52	9.87	1.04	0.0789
COM	499.92	10.15	500.36	9.78	1.08	0.1218	499.16	10.39	500.84	9.42	1.22	0.0853
PIC	502.62	8.50	496.43	10.73	1.60	0.1482	502.92	8.67	497.19	10.54	1.48	0.0939
BLD	503.98	8.82	494.72	8.99	1.04	0.1197	503.65	12.39	494.98	8.45	2.15	0.1133
MR	503.98	8.06	495.54	10.27	1.63	0.1496	503.99	8.44	495.89	9.72	1.33	0.0890
PA	503.64	8.66	495.20	9.83	1.29	0.1332	505.37	8.69	494.51	9.16	1.11	0.0815
ARI	501.03	9.91	498.72	10.08	1.03	0.1193	500.67	10.51	499.13	9.82	1.14	0.0826
DS	501.77	9.71	497.70	10.03	1.07	0.1212	501.76	11.50	497.55	9.29	1.53	0.0956
DAS	500.36	15.09	492.79	14.19	1.13	0.1247	505.38	8.65	494.54	8.15	1.13	0.0820
SYS	504.00	8.69	494.63	9.12	1.10	0.1232	505.38	9.96	493.42	8.88	1.26	0.0867
Average	501.93	6.30	497.02	7.48	1.11	0.1236	502.28	7.33	497.41	6.53	1.26	0.0866
N	170		127				338		332			

Table 3.3. The *g* loadings across age groups with WAIS-III normative samples

Subtest*	USA Age 16-44	Age 45-89		Canada Age 16-44	Age 44-80		Australia Age 16-44	Age 45-89		Netherlands Age 16-44	Age 45-86	
VOC	0.8329	0.8272	–	0.8373	0.7570	–	0.7681	0.8006	+	0.8175	0.8065	–
SIM	0.8069	0.8409	+	0.8358	0.8271	–	0.8197	0.8272	+	0.8239	0.8282	+
INF	0.7806	0.8011	+	0.7614	0.7498	–	0.7316	0.7323	+	0.8177	0.7853	–
COM	0.7759	0.8109	+	0.7999	0.7670	–	0.7740	0.8112	+	0.8175	0.7881	–
PIC	0.6562	0.7486	+	0.5659	0.7003	+	0.4939	0.7046	+	0.6491	0.5735	–
BLD	0.7274	0.7459	+	0.6060	0.7578	+	0.6918	0.7488	+	0.5518	0.5362	–
MR	0.7273	0.7586	+	0.6619	0.7778	+	0.6635	0.7524	+	0.7427	0.6583	–
PA	0.6785	0.7506	+	0.5457	0.7060	+	0.6012	0.7257	+	0.6824	0.5409	–
ARI	0.7785	0.7535	–	0.7542	0.7812	+	0.7408	0.6878	–	0.7714	0.6527	–
DS	0.6007	0.5729	–	0.4402	0.5180	+	0.5501	0.6475	+	0.6001	0.5472	–
DAS	0.5930	0.7232	+	0.3827	0.5682	+	0.2063	0.4533	+	0.6645	0.6001	–
SYS	0.6874	0.7853	+	0.4974	0.7048	+	0.5386	0.6231	+	0.6388	0.6079	–
Eigenvalue of the PCI	6.30	6.98	*F* = 1.11	5.20	6.27	*F* = 1.21	5.10	6.16	*F* = 1.21	6.23	5.37	*F* = .86
Average inter-correlation(γ)	.48	.54	*d* = .0426	.38	.48	*d* = .0673	.37	.47	*d* = .1290	.48	.40	*d* = .0717
N	1195	1255		652	453		170	127		338	332	

Note: *VOC = Vocabulary, SIM = Similarity, INF = Information, COM = Comprehension, PIC = Picture completion, BLD = Block design, MR = Matrix reason, PA = Picture arrangement, ARI = Arithmetic, DS = Digit span, DAS = Digit symbol-coding, SYS = Symbol search. +: *g* loadings increased from young to old group. –: *g* loadings decreased from young to old group.

From Table 3.3, the *g* loading of the PIC, BLD, MR, PA, DAS, and SYS subtests did not decrease as a function of age within the US, Canada, and Australia normative samples. The *g* loading of the SIM, INF, and COM subtests in the older age group was higher than the *g* loading for the young age group only for the US and Australian normative samples. For the whole test battery, Table 3.3 also shows that the average intercorrelation (γ) is increased from the young age group to the older age group for all but the Netherlands normative sample. The Cohen's *d* effect sizes (Hays, 1994, p. 411) are small, ranging from .0426 to .1290. These findings would suggest that there was no compelling evidence to support SLODR. However, for the Netherlands normative sample, the *g* loadings are higher for the older age group for all subtests except SIM.

g Loadings Across Ability Groups

In the ability group study, the WAIS-III normative sample was divided into a low ability and a high ability group based on the average CTT ability scores. Correlations between age and ability for overall sample and by ability group are reported in Table 4.1. As expected, the correlations didn't show any large changes across ability groups.

In order to assure equal variances across groups, the cutoff points were 498, 495, 500, and 497 for US, Canadian, Australian, and Netherlands samples, respectively. The WAIS-III CTT-ability score mean and *SD* between the lower and the higher ability groups are presented in Table 4.2. The *F* testing for the average variance indicated that there are no significant differences between two group variances ($p > .10$). For each subtest, although the *F* testing results

Table 4.1. Correlations between age and ability scores for WAIS-III normative sample, overall, and by ability group

	USA			Canada			Australia			Netherlands		
Subtest	Overall	Low	High	overall	Low	High	overall	Low	High	overall	Low	High
VOC	0.06	0.37	0.39	–0.03	0.36	0.22	0.05	0.23	0.43	0.17	0.42	0.40
SIM	–0.21	0.03	0.08	–0.28	–0.04	–0.01	–0.27	–0.11	0.03	–0.10	0.21	0.03
INF	0.05	0.32	0.36	–0.01	0.30	0.23	0.04	0.23	0.37	0.06	0.36	0.26
COM	0.02	0.30	0.36	–0.09	0.24	0.17	0.05	0.25	0.38	0.09	0.39	0.27
PIC	–0.42	–0.29	–0.23	–0.38	–0.18	–0.18	–0.35	–0.34	0.04	–0.37	–0.21	–0.34
BLD	–0.48	–0.23	–0.39	–0.49	–0.32	–0.34	–0.54	–0.45	–0.41	–0.43	–0.07	–0.47
MR	–0.55	–0.43	–0.44	–0.51	–0.47	–0.30	–0.51	–0.44	–0.30	–0.47	–0.35	–0.42
PA	–0.57	–0.51	–0.42	–0.55	–0.49	–0.40	–0.49	–0.45	–0.22	–0.56	–0.36	–0.60
ARI	–0.13	0.21	0.10	–0.18	0.15	0.06	–0.15	0.00	0.18	–0.04	0.37	0.05
DS	–0.25	0.02	–0.14	–0.25	–0.02	–0.10	–0.25	–0.12	–0.02	–0.18	0.13	–0.16
DAS	–0.65	–0.56	–0.57	–0.54	–0.45	–0.42	–0.30	–0.21	–0.14	–0.56	–0.35	–0.59
SYS	–0.62	–0.51	–0.56	–0.62	–0.57	–0.50	–0.55	–0.49	–0.40	–0.57	–0.36	–0.61
N	2450	894	1556	1105	246	859	297	135	162	670	212	458

Table 4.2. The CTT-Ability score means and *SD* across ability groups with WAIS-III normative samples

	USA						Canada					
	Low		High				Low		High			
Subtest[1]	Mean	*SD*	Mean	*S*	F[2]	d[3]	Mean	*SD*	Mean	*SD*	*F*	*d*
VOC	492.07	8.30	504.52	7.82	1.13	0.0446	489.77	9.16	502.91	8.21	1.25	0.0468
SIM	491.09	8.18	505.14	7.00	1.37	0.0491	487.73	8.81	503.48	7.03	1.57	0.0526
INF	492.46	7.52	504.35	8.49	1.28	0.0474	490.54	8.89	502.76	8.60	1.07	0.0434
COM	492.04	8.75	504.52	7.46	1.38	0.0493	489.70	8.93	502.98	8.14	1.20	0.0461
PIC	491.86	10.72	504.68	5.65	3.60	0.0797	488.44	11.69	503.39	6.30	3.44	0.0779
BLD	491.75	6.58	504.80	8.31	1.59	0.0530	488.85	6.94	503.16	8.31	1.43	0.0502
MR	491.19	6.64	505.14	7.74	1.36	0.0489	487.88	7.28	503.50	7.65	1.10	0.0441
PA	491.55	8.13	504.87	7.34	1.22	0.0464	489.10	9.00	503.05	7.79	1.33	0.0485
ARI	491.97	7.09	504.61	8.47	1.42	0.0501	489.24	7.39	503.13	8.41	1.30	0.0478
DS	493.43	8.07	503.86	9.00	1.25	0.0468	491.47	8.49	502.32	9.20	1.17	0.0455
DAS	492.40	8.70	504.38	7.88	1.22	0.0464	489.47	10.17	502.32	9.76	1.09	0.0437
SYS	491.86	7.94	504.72	7.85	1.02	0.0425	489.45	7.81	503.22	8.16	1.09	0.0438
Average[4]	491.97	5.03	504.63	4.15	1.09	0.0437	489.30	4.51	503.02	4.48	1.16	0.0452
N	894		1556				246		859			

Note: [1]VOC = Vocabulary, SIM = Similarity, INF = Information, COM = Comprehension, PIC = Picture completion, BLD = Block design, MR = Matrix reason, PA = Picture arrangement, ARI = Arithmetic, DS = Digit span, DAS = Digit symbol-coding, SYS = Symbol search. [2]*F* testing for each subtest may be misleading because very little difference will be identified as significant difference due to big sample size. [3]*d* is effect size. [4]*F* variance ratio for the mean of the eight subtest variances is calculated using the following method (Hartmann & Teasdale, 2005, p. 1197):

$F = SD^2_{Average(Larger)}/SD^2_{Average(Smaller)}$, where

$SD^2_{Average} = (SD^2_{VOC} + SD^2_{SIM} + SD^2_{INF} + SD^2_{COM} + SD^2_{PIC} + SD^2_{BLD} + SD^2_{MR} + SD^2_{PA} + SD^2_{ARI} + SD^2_{DS} + SD^2_{DAS} + SD^{2)}_{SYS}$ and *F* ratio for the average is 1.09, 1.16, 1.56 and 1.40 for USA, Canada, Australia, and Netherlands sample, respectively ($p > .10$, two tailed).

Table 4.2. (continued)

	Australia						Netherlands					
	Low		High				Low		High			
Subtest	Mean	*SD*	Mean	*SD*	*F*	*d*	Mean	*SD*	Mean	*SD*	*F*	*d*
VOC	494.25	9.58	504.83	7.45	1.65	0.0540	491.96	9.98	503.77	7.52	1.76	0.0557
SIM	493.40	9.23	505.95	6.15	2.25	0.0630	490.10	10.17	504.48	6.24	2.66	0.0684
INF	494.32	9.33	504.73	7.91	1.39	0.0495	491.18	7.84	504.08	8.05	1.06	0.0431
COM	494.33	9.60	504.93	7.42	1.68	0.0543	491.48	10.70	503.93	6.55	2.66	0.0685
PIC	494.59	10.79	504.46	6.46	2.79	0.0701	492.42	11.70	503.63	6.70	3.04	0.0732
BLD	493.34	8.52	505.58	7.39	1.33	0.0484	490.43	9.06	503.49	10.03	1.22	0.0464
MR	493.26	9.21	506.30	5.87	2.46	0.0658	491.01	9.37	504.12	7.08	1.75	0.0555
PA	493.44	9.28	505.51	6.96	1.78	0.0560	491.43	9.03	503.94	8.51	1.13	0.0446
ARI	493.87	9.09	505.19	7.59	1.44	0.0503	491.18	8.10	503.95	8.38	1.07	0.0435
DS	493.96	7.20	505.09	9.24	1.64	0.0538	492.48	8.99	503.00	9.70	1.16	0.0453
DAS	490.28	15.32	502.83	12.46	1.51	0.0516	492.27	8.79	503.59	8.37	1.10	0.0441
SYS	494.11	9.20	504.90	7.77	1.40	0.0497	491.21	10.80	503.27	9.09	1.41	0.0499
Average	493.60	5.46	505.02	3.45	1.56	0.0525	491.43	5.27	503.77	4.26	1.40	0.0497
N	135		162				212		458			

Table 4.3. The *g* loadings across ability groups with WAIS-III normative samples

	USA			Canada			Australia			Netherlands		
Subtest*	Low	High		Low	High		Low	High		Low	High	
VOC	0.5805	0.6223	+	0.6236	0.7359	+	0.6602	0.7775	+	0.7718	0.4230	–
SIM	0.6867	0.6383	–	0.6737	0.7476	+	0.7533	0.6114	–	0.7225	0.5497	–
INF	0.4927	0.5876	+	0.5807	0.6493	+	0.5702	0.6409	+	0.6293	0.5137	–
COM	0.5300	0.5456	+	0.6721	0.6956	+	0.6608	0.7695	+	0.7251	0.4206	–
PIC	0.6479	0.3812	–	0.3863	0.3568	–	0.6184	0.0373	–	0.4946	0.3250	–
BLD	0.6234	0.5121	–	0.3178	0.4843	+	0.6614	0.0627	–	0.0344	0.5456	+
MR	0.5301	0.5101	–	0.3872	0.5157	+	0.5446	0.0774	–	0.4085	0.5543	+
PA	0.5657	0.4605	–	0.4216	0.3898	–	0.5185	0.1569	–	0.2115	0.5475	+
ARI	0.5295	0.5709	+	0.5266	0.6075	+	0.5623	0.3344	–	0.4761	0.5240	+
DS	0.5091	0.3199	–	0.2866	0.2462	–	0.4454	0.2637	–	0.5609	0.3673	–
DAS	0.6727	0.3037	–	0.2613	0.2662	+	0.0075	0.0040	–	0.4421	0.5246	+
SYS	0.7247	0.4023	–	0.3727	0.3194	–	0.4855	0.0318	–	0.4888	0.5165	+
Eigenvalue of the PC1	4.26	3.00	*F* = .70	2.78	3.39	*F* = 1.22	3.90	2.20	*F* = .56	3.47	2.88	*F* = .83
Average inter-correlation(γ)	.30	.18	*d* = .0351	.16	.22	*d* = .0799	.26	.11	*d* = .0872	.22	.17	*d* = .0757
N	894	1556		246	859		135	162		212	458	

Note: *VOC = Vocabulary, SIM = Similarity, INF = Information, COM = Comprehension, PIC = Picture completion, BLD = Block design, MR = Matrix reason, PA = Picture arrangement, ARI = Arithmetic, DS = Digit span, DAS =, Digit symbol-coding, SYS = Symbol search. +: *g* loadings increased from lower to higher ability group. –: *g* loadings decreased from lower to higher ability group.

can be ignored because of the large sample sizes, *F* ratios for subtest PIC were relatively higher across the four countries. These relatively higher *F* values resulted in a decrease of the *g* loading from the lower to the higher ability groups across the four countries. However, the Cohen's *d* effect sizes (Hays, 1994, p. 411) of ability score variance differences between the two ability groups for the average and subtests were still small, ranging from 0.0405 to 0.0797. These results showed that the equal variance assumption was still met in this study.

The *g* loadings across ability groups for the four WAIS-III normative samples are presented in Table 4.3.

From Table 4.3, the *g* loadings of the VOC, INF, and COM subtests did not decrease as a function of ability with the US, Canadian, and Australian normative samples. The same results can also be seen for the BLD, MR, PA, ARI, DAS, and SYS subtests in the Netherlands normative sample. These findings again do not support SLODR. However the *g* loading for PIC and DS was lower for the higher ability group across the US, Canadian, Australian, and Netherlands normative samples. This result can be also found in BLD, MR, and PA for both the US and Australian normative samples. For the whole test battery, Table 4.3 shows that the average intercorrelation (γ) decreased from the lower ability group to the higher ability groups for the US, Australian, and Netherlands, but not for the Canadian normative samples. The Cohen's *d* effect sizes (Hays, 1994, p. 411) are small, ranging from .0351 to .0872. This result again does not support Spearman's hypothesis that the *g* saturation of cognitive ability tests decreases as a function of ability.

Discussion

From the WISC-IV study, most of the subtest and full battery results did not support SLODR (by ability level or by age). However, the WAIS-III study provides only partial evidence against SLODR. Two points need to be discussed here. First, the *g* loading has a strong relationship with both the sample size and score variability even if the score variances are equally controlled by classic methods. Some studies have demonstrated that the higher the variance, the higher the *g* loading for the subtest (Hartmann & Teasdale, 2004). Second, the weaker evidence against SLODR from the WAIS-III results may be because the variability of ability scores for adults is smaller than that for children, and the direction of ability development differs between children and adults. Children's cognitive abilities increase with age in contrast to adult cognitive ability that appears somewhat more variable depending on the factor being assessed and at what period in time (e.g., decreases in processing speed, especially from the middle years onward, vs. more stable scores for crystallized intelligence). Further, these standardization studies excluded individuals manifesting a variety of conditions ranging from uncorrected hearing or visual impairment and problems with thinking, to medical and psychiatric conditions that may impact cognitive functioning (e.g., stroke, Parkinson's disease, Alzheimer's dementia). With aging, there is an increasing likelihood of such conditions affecting a larger portion of the population and these groups are not included in the standardization studies (except as "clinical groups"). Also, there is some evidence to suggest that low ability adults tend to die at an earlier age (Whalley & Deary, 2001). All of these factors may certainly create an even greater restriction in range of ability in older adults and, thus, provide results that neither support or refute SLODR.

This study did not support either version of SLODR (by ability level or by age) using large normative samples of children for the WISC-IV from the United States, Canada, and Australia, and adults for the WAIS-III from the same three countries as well as from The Netherlands. The Wechsler scales are often-used measures for assessing intellectual ability across a wide age and ability range and certainly tap both crystallized and fluid abilities. Using the WISC-IV and WAIS-III provided data across an 80-year age span. These large standardization studies not only contribute a substantial data set for studies of the kind reported here but the tests have been carefully adapted for use in other English and non-English-speaking countries such as the Netherlands. Thus, these tests, with their *g* saturation, and the standardization data sets reported here provide a robust and compelling test of SLODR.

As regards ability level effects, the null findings here are somewhat surprising. Previous positive findings across a range of test batteries (Abad et al., 2003; Deary et al, 1996; Detterman & Daniel, 1989; Der & Deary, 2003; Legree et al., 1996; Jensen, 2003; te Nijenhuis & Hartmann, 2006) outnumber negative ones (Facon, 2004; Fogarty & Stankov, 1995; Hartmann & Reuter, 2006; Hartmann & Teasdale, 2004, 2005). In particular, SLODR by ability level has been found in studies using earlier versions of the WAIS and WISC (Detterman & Daniel, 1989; Lynn, 1992) and in a recent study using the Spanish version of the WAIS-III (Abad et al, 2003) based on age-corrected standard scores. Similar results can be found using WISC-IV data based on age-corrected standard scores. It has also been noted that batteries of diverse content such as the Wechsler tests should be the most likely candidates to show SLODR (Hartmann & Teasdale, 2005).

The null findings for SLODR by age are perhaps less surprising, since this result is consistent with those that have emerged from the majority of recent large-scale studies (Carroll, 1993; Juan-Espinosa et al., 2000; Rietveld et al., 2003).

It is possible that differences in the internal reliability of subtests between groups could affect the findings reported here, since lower reliability depresses test-test correlations and *g*-loadings (Hartmann & Teasdale, 2005; Jensen, 2003). In the case of the present null findings, the observation of SLODR might be suppressed if reliabilities were much lower in the low ability or young groups. Item-level data were available for the WISC but not for the WAIS samples. The internal reliability (Cronbach's α) of the WISC-IV subtests by age group and ability group are presented in Table 5.1 and Table 5.2.

It can be seen that there is no indication of a systematic trend in reliability with either age or ability, with the numbers of cases of a slightly lower or higher subtest internal reliability in the older age or higher ability groups being near equal and with all absolute differences in internal reliability being very small. The effect size *d* (Hays, 1994, p. 411) ranged from 0.0481 to 0.0812 for ability groups, and ranged from 0.0523 to 0.0829 for age groups. These results indicate that the nonobservation of SLODR for the WISC is not related to differential subtest reliability.

One view of the present results is that the existence or nonexistence of SLODR would be best pursued by adopt-

Table 5.1. Internal reliability (α) of the WISC-IV subtests by age group

	US				Canada				Australia			
Subtest	Age6-8	Age9-16	*F*	*d*	Age6-8	Age9-16	*F*	*d*	Age6-8	Age9-16	*F*	*d*
SI	0.85	0.88	1.04	0.0487	0.84	0.87	1.04	0.0689	0.86	0.86	1.00	0.0767
VC	0.85	0.90	1.06	0.0493	0.84	0.89	1.06	0.0697	0.81	0.88	1.09	0.0799
CO	0.78	0.85	1.09	0.0500	0.78	0.84	1.08	0.0703	0.75	0.74	1.01	0.0772
BD	0.79	0.83	1.05	0.0491	0.76	0.80	1.05	0.0695	0.76	0.81	1.07	0.0792
PS	0.84	0.79	1.06	0.0494	0.84	0.79	1.06	0.0698	0.83	0.74	1.12	0.0812
MR	0.89	0.88	1.01	0.0481	0.88	0.86	1.02	0.0685	0.89	0.88	1.01	0.0771
DS	0.70	0.78	1.11	0.0505	0.69	0.79	1.14	0.0724	0.69	0.76	1.10	0.0805
LN	0.79	0.78	1.01	0.0482	0.80	0.74	1.08	0.0704	0.79	0.74	1.07	0.0792
N	600	1600			300	800			235	616		

Table 5.2. Internal reliability (α) of the WISC-IV subtests by ability group

	US				Canada				Australia			
Subtest	Low	High	*F*	*d*	Low	High	*F*	*d*	Low	High	*F*	*d*
SI	0.79	0.85	1.08	0.0503	0.79	0.85	1.08	0.0715	0.84	0.84	1.00	0.0769
VC	0.80	0.88	1.10	0.0508	0.80	0.87	1.09	0.0719	0.77	0.87	1.13	0.0817
CO	0.76	0.83	1.09	0.0507	0.77	0.82	1.06	0.0712	0.72	0.72	1.00	0.0769
BD	0.74	0.81	1.09	0.0507	0.72	0.79	1.10	0.0723	0.69	0.78	1.13	0.0817
PS	0.82	0.74	1.11	0.0510	0.80	0.74	1.08	0.0717	0.77	0.69	1.12	0.0812
MR	0.86	0.84	1.02	0.0490	0.86	0.84	1.02	0.0698	0.86	0.84	1.02	0.0778
DS	0.67	0.75	1.12	0.0513	0.64	0.77	1.20	0.0757	0.65	0.74	1.14	0.0820
LN	0.75	0.74	1.01	0.0488	0.77	0.66	1.17	0.0745	0.78	0.67	1.16	0.0829
N	577	1623			283	817			233	618		

ing a meta-analytic approach to the growing body of findings in this area. Another perspective is that the theoretical basis of SLODR involves a cognitive-developmental perspective. Therefore, particularly in the case of putative age effects, a longitudinal approach in which the evolution over time of test *g*-loadings for the same group is examined might provide a useful approach to this issue.

References

Abad, F.J., Colom, R., Juan-Espinosa, M., & García, L.F. (2003). Intelligence differentiation in adult samples. *Intelligence, 31,* 157–166.

Anderson, M. (1992). *Intelligence and development. A cognitive theory.* Oxford: Blackwell.

Bailey, M.J., & Revelle, W. (1991). Increased heritability for lower IQ levels? *Behavior Genetics, 21,* 397–204.

Bates, T. (2007). Fluctuating asymmetry and intelligence. *Intelligence, 35,* 41–46.

Bayley, N. (1955). On the growth of intelligence. *American Psychologist, 10,* 805–818.

Burt, C. (1954). The differentiation of intellectual ability. *British Journal of Educational Psychology, 24,* 76–90.

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, UK: Cambridge University Press.

Cattell, R.B. (1987). *Intelligence: Its structure, growth, and action.* Amsterdam: North Holland.

Cherny, S.S., Cardon, L.R., Fulker, D.W., & DeFries, J.C. (1992). Differential heritability levels across levels of cognitive ability. *Behavior Genetics, 22,* 153–162.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Colom, R., Abad, F.J., Garcia, L.F., & Juan-Espinosa, M. (2002). Education, Wechsler's full scale IQ, and *g*. *Intelligence, 30,* 449–462.

Deary, I.J., Egan, V., Gibson, G.J., Austin, E.J., Brand, C.R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence, 23,* 105–132.

Der, G., & Deary, I.J. (2003). IQ, reaction time, and the differentiation hypothesis. *Intelligence, 31,* 491–503.

Detterman, D.K., & Daniel, M.H. (1989). Correlations of mental tests with each other and with cognitive variables are highest in low-IQ groups. *Intelligence, 13,* 349–359.

Detterman, D.K., Thompson, L.A., & Plomin, R. (1990). Differences in heritability across groups differing in ability. *Behavior Genetics, 20,* 369–384.

Facon, B. (2004). Are correlations between cognitive abilities highest in low-IQ groups during childhood? *Intelligence, 32,* 391–401.

Fogarty, G.J., & Stankov, L. (1995). Challenging the "law of diminishing returns." *Intelligence, 21,* 157–174.

Furlow, F.B., Armijo-Prewitt, T., Gangestad, S.W., & Thornhill, R. (1997). Fluctuating asymmetry and psychometric intelligence. *Proceedings of the Royal Society of London. Series B, Biological Sciences, 264,* 823–829.

Garrett, H.E. (1946). A developmental theory of intelligence. *American Psychologist, 1,* 372–378.

Georgas, J., Weiss, L.G., van de Vijver, F.J.R., & Saklofske, D.H. (2003). *Culture and children's intelligence: Cross-cultural analysis of the WISC-III.* San Diego: Academic Press.

Ghisletta, P., & Lindenberger, U. (2003). Age-based structural dynamics between perceptual speed and knowledge in the Berlin Aging Study: Direct evidence for ability dedifferentiation in old age. *Psychology and Aging, 18,* 696–713.

Hartmann, P. (2006). Spearman's law of diminishing returns. A look at age differentiation. *Journal of Individual Differences, 27,* 199–207.

Hartmann, P., & Nyborg, H. (2004). Spearman's "law of diminishing returns": A critical eye on a century of methods, results, and current standing of the theory. In P. Hartmann (Ed.), *Investigating Spearman's "law of diminishing returns"* [Ph.D. Thesis, 2006] (pp. 31–190). Aarhus: Aarhus University.

Hartmann, P., & Reuter, M.(2006). Spearman's "law of diminishing returns" tested with two methods. *Intelligence, 34*, 47–62.

Hartmann, P., & Teasdale, T.W. (2004). A test of Spearman's "law of diminishing returns" in two large samples of Danish military draftees. *Intelligence, 32,* 499–508.

Hartmann, P., & Teasdale, T.W. (2005). Spearman's "law of diminishing returns" and the role of test reliability in a large sample of Danish military draftees. *Personality and Individual Differences, 39,* 1193–1203.

Hays, L.W. (1994). *Statistics*. Orlando, FL: Harcourt..

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

Jensen, A.R. (1997). The puzzle of nongenetic variance. In R.J. Sternberg & E.L. Grigorenko (Eds.), *Heredity, intelligence, and environment* (pp. 42–88). Cambridge: Cambridge University Press.

Jensen, A.R. (2003). Regularities in Spearman's law of diminishing returns. *Intelligence, 31,* 95–105.

Jensen, A.R., & Weng, L-J. (1994). What is good *g? Intelligence, 18,* 231–258.

Juan-Espinosa, M., García, L.F., Colom, R., & Abad, F.J. (2000). Testing the age-related differentiation hypothesis through the Wechsler's scales. *Personality and Individual Differences, 29,* 1069–1075.

Kaiser, H.F. (1968). A measure of the average intercorrelation. *Educational and Psychological Measurement, 28*, 245–247.

Kaufman, A.S., Flanagan, D.P., Alfonso, V.C., & Mascolo, J.T. (2006). Test review: Wechsler Intelligence Scale for Children, 4th ed. (WISC-IV). *Journal of Psychoeducational Assessment, 24*, 278–295.

Legree, P.J., Pifer, M.E., & Grafton, F.C. (1996). Correlations among cognitive abilities are lower for higher ability groups. *Intelligence, 23,* 45–57.

Lynn, R. (1992). Does Spearman's *g* decline at high levels? Some evidence from Scotland. *Journal of Genetic Psychology, 153,* 229–230.

Miller, G. (2000). Mental traits as fitness indicators: Expanding evolutionary psychology's adaptionism. *Annals of the New York Academy of Sciences, 907,* 62–74.

Prokosch, M.D., Yeo, R.A., & Miller, G.F. (2005). Intelligence test with higher *g*-loading show higher correlations with body symmetry: Evidence for a general fitness factor mediated by developmental stability. *Intelligence, 33,* 203–213.

Rietveld, M.J.H., Dolan, C.V., van Baal, G.C.M., & Boomsma, D.I. (2003). A twin study of differentiation of cognitive abilities in childhood. *Behavior Genetics, 33,* 367–381.

Sattler, J. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego, CA: Author.

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* London: Macmillan.

te Nijenhuis, J., & Hartmann, P. (2006). Spearman's "law of diminishing returns" in samples of Netherlands and immigrant children and adults. *Intelligence, 34,* 437–447.

Tideman, E., & Gustafsson, J. (2004). Age-related differentiation of cognitive abilities in ages 3–7. *Personality and Individual Differences, 36,* 1965–1974.

Tulsky, D.S., Saklofske, D.H., Chelune, G.J., Ivnik, R.J., Prifitera, A., Heaton, R.K. et al. (2003). *Clinical interpretation of the WAIS-III and WMS-III.* San Diego: Academic Press.

Wechsler, D. (1997). *Manual for the Wechsler Adult Intelligence Scale (3rd ed.).* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003). *Manual for the Wechsler Intelligence Scale for Children (4th ed.).* San Antonio, TX: The Psychological Corporation.

Whalley, L.J., & Deary, I.J. (2001). Longitudinal cohort study of childhood IQ and survival up to age 76. *British Medical Journal, 322,* 1–5.

Donald H. Saklofske

Division of Applied Psychology
University of Calgary
2500 University Drive, N.W.
Calgary, Alberta T2N 1N4
Canada
Tel. +1 403 220-4050
E-mail don.saklofske@ucalgary.ca