Practice effects on the WAIS-III across 3- and 6-month Intervals

Article in The Clinical Neuropsychologist · March 2002
DOI: 10.1076/clin.16.1.57.8329 · Source: PubMed

4 authors, including:

Michael R Basso
University of Tulsa
77 PUBLICATIONS 1,878 CITATIONS
SEE PROFILE

Bradley Axelrod
U.S. Department of Veterans Affairs
225 PUBLICATIONS 3,019 CITATIONS
SEE PROFILE

Available from: Bradley Axelrod
Retrieved on: 18 September 2016
Practice Effects on the WAIS-III Across 3- and 6-Month Intervals

Michael R. Basso1,2, Francine D. Carona1, Natasha Lowery1, and Bradley N. Axelrod3
1University of Tulsa, Tulsa, OK, USA, 2University of Oklahoma Schusterman Health Sciences Center, Tulsa, OK, USA, and 3John D. Dingell Department of Veterans Affairs Medical Center, Detroit, MI, USA

ABSTRACT

Fifty-one participants (age $M = 24.6$; education $M = 14.4$ years) were administered the Wechsler Adult Intelligence Scale – Third Edition (WAIS-III) at baseline and at an interval of either 3 or 6 months later. Full Scale IQ (FSIQ), Verbal IQ (VIQ), Performance IQ (PIQ), Verbal Comprehension Index (VCI), Perceptual Organization Index (POI), and Processing Speed Index (PSI) scores improved significantly across time, whereas no significant change occurred on the Working Memory Index. Specifically, test scores increased approximately 3, 11, 6, 4, 8, and 7 points, respectively on the VIQ, PIQ, FSIQ, VCI, POI, and PSI for both groups. Notably, the degree of improvement was similar regardless of whether the inter-test interval was 3 or 6 months. These findings suggest that prior exposure to the WAIS-III yields considerable increases in test scores. Reliable change indices indicated that large confidence intervals might be expected. As such, users of the WAIS-III should interpret reevaluations across these intervals cautiously.

Many applications of neuropsychological testing involve repeated examinations across time. For instance, studies of treatment efficacy frequently employ serial assessment of neuropsychological function to determine the utility of interventions. In clinical practice, patients are often reexamined to determine whether deterioration or improvement has occurred. In these scientific and clinical contexts, performance variations bear important implications, including the prediction of disease course, generation of treatment recommendations, and monitoring of rehabilitation progress (Lezak, 1995).

Despite its significance, accurate measurement of change may be confounded by the effect of practice. Specifically, after exposure to an initial assessment, an individual may refine test-taking strategies, thereby enhancing subsequent performances (cf. Basso, Bornstein, & Lang, 1999).

Additionally, because test items are no longer novel, responses may become more automatic and demand less attention and concentration (cf. Lezak, 1995). Consequently, abilities assessed upon reexamination may differ from those assessed during an initial evaluation.

Among the measures presumed most susceptible to practice effects are those that involve speed, demand an infrequently practiced response, or have a single and memorable solution (Lezak, 1995; McCaffrey, Ortega, Orsillo, Nelles, & Haase, 1992). Several subtests comprising the Wechsler Adult Intelligence Scale – Revised (WAIS–R; Wechsler, 1981) possess these characteristics, and are noted for yielding gains with retesting (Kaufman & Lichtenberger, 1999). For instance, normal individuals and clinical populations typically display gains of about 3 points in Verbal IQ (VIQ), 9 points in Performance IQ (PIQ)
and 6 points in Full Scale IQ (FSIQ) (Kaufman, 1990).

While considerable research exists concerning practice effects on the WAIS–R, relatively little is known about the Wechsler Adult Intelligence Scale – Third Edition (WAIS-III; Wechsler, 1997). One of the few studies to address retest effects on the WAIS-III appears in the WAIS-III–WMS-III Technical Manual (Tulsky & Zhu, 1997). In this investigation, the WAIS-III was readministered once to a group of approximately 400 individuals across a mean retest interval of 34.6 days. Across four age groups (ages 16–29; 30–54; 55–74; & 75–89), average increases on the VIQs, PIQs, and FSIQs ranged from 2.5–3.2, 2.5–8.3, and 2.0–3.2 points on the three IQ scores, respectively. These increases were statistically significant, and were similar to those observed in research involving the WAIS–R (cf. Kaufman, 1990). In addition, the investigation detailed the rates of improvement on WAIS-III Index Scores. These Index Scores are an innovation over the WAIS–R, and reflect factor-analytically derived scores. For the Verbal Comprehension, Perceptual Organization, Working Memory, and Processing Speed Indexes, average increments ranged from 1.9–3.2, 2.7–7.3, 1.3–3.1, and 1.3–6.0 points, respectively. Consequently, meaningful increases in WAIS-III IQs and Index Scores appear to occur over relatively short time intervals.

A factor that bears investigation is the length of the test-retest interval. Tulsky and Zhu (1997) examined performance increments over a relatively short time interval (approximately 35 days). While they found significant improvement, such score increments may not extend to longer interest intervals. For example, across longer retest intervals, individuals may remember fewer details of testing, thereby minimizing the practice effects on WAIS-III scores. Furthermore, the short interest interval assessed by past research may be irrelevant to clinicians and researchers who tend to reexamine neuropsychological function over longer intervals such as 3, 6, or 12 months. As yet, however, the degree to which practice influences performance on the WAIS-III over such lengthy inter-test intervals remains uncertain.

To address these issues, the present study examined the effects of prior testing on subsequent WAIS-III performance. The WAIS-III was administered at baseline to two groups of individuals. One group was reevaluated 3 months later, whereas another group was reexamined 6 months later. Neurologically normal individuals participated rather than a clinical sample. This was done to avoid potential confounds associated with genuine improvement. For example, in neurological patients, changes across time may reflect recovery from illness or injury. By contrast, in a neurologically normal sample, any increase in test scores would be exclusively attributable to practice effects.

METHOD

Participants

During each assessment interval, participants were screened for neurological and psychiatric disease, head injury, learning disabilities, and other medical illness. Seven potential participants were excluded after meeting at least one of these criteria. The resulting group included 73 men and women who were recruited from the community. The group was composed of 64 Caucasians, 2 Native Americans, 3 Asian Americans, 1 African American, and 1 Hispanic. Two were of unreported ethnicity. At baseline, average age of the entire sample was 23.8 years ($SD = 7.9$), and mean education level was 14.5 years ($SD = 1.5$).

Of the original 73 participants, 51 returned for the follow-up assessment. To rule out possible bias between those who completed the study and those who did not, significant differences between baseline performance as well as demographic variables were tested. Univariate analyses of variance (ANOVAs) were computed for age, education, and each WAIS-III Index. As shown in Table 1, there were no significant differences in IQ scores between the two groups. With respect to Index Scores, the groups differed only on Processing Speed Index, $F(1, 67) = 4.80, p < .05$. Those who completed the study had a mean Processing Speed Index of 109.3 ($SD = 12.9$), whereas those who did not finish the study had a mean score of 116.6 ($SD = 13.3$). This finding notwithstanding, these findings suggest that those who completed the retesting differed little from those who did not.

Of the 51 who completed the two assessment intervals, participants were randomly assigned to be
retested either 3 or 6 months later. The 3-month retest group consisted of 17 women and 6 men, whereas the 6-month retest group included 18 women and 10 men. The 3-month group was comprised of 20 Caucasians, 2 Asian Americans, and 1 Native American. The 6-month group involved 25 Caucasians, 1 Native American, 1 Hispanic, and 1 unreported ethnicity. Nonparametric tests showed the groups did not differ according to these characteristics. Regarding education level, those retested at 3 months averaged 20.3 years (SD = 2.3 years, range = 12–18 years), whereas those reevaluated 6 months later averaged 17.7 years (SD = 1.4 years, range = 13–19 years). An F-test revealed no significant difference according to this characteristic. With respect to age, the 3-month group averaged 20.3 years (SD = 2.3 years, range = 17–29 years). The 6-month group had an average age of 28.0 years (SD = 10.4 years, range = 19–53 years). Accordingly, the two groups differed significantly according to age, $F(1, 49) = 12.03$, $p = .001$.

Materials and Procedures

At baseline and follow-up, participants were administered the WAIS-III (Wechsler, 1997). Clinical psychology doctoral students administered the WAIS-III in accordance with the test’s standardization rules. Thus, all participants received the same test instructions, and they were requested to give their best performance across all tasks.

Four examiners administered the tests. Prior to their involvement, each had completed at least 2 years of graduate training in clinical psychology, and had completed coursework in intellectual assessment. Moreover, they had at least 2 years of supervised clinical assessment experience. Their WAIS-III administration and scoring acumen had been repeatedly monitored and evaluated during the course of their training, and they were judged proficient by instructors and clinical supervisors. Additionally, to ensure adherence to standardization instructions, the doctoral students were evaluated by the first author (an Associate Professor of Psychology and a licensed Clinical Neuropsychologist) prior to their participation in the study. Scoring was completed by a single student who had completed 3 years of didactic and practical training. This student checked each protocol twice to further increase accuracy. Owing to logistical constraints, the same examiner may not have readministered the WAIS-III to individual participants.1

RESULTS

To evaluate whether the 3-month and 6-month groups displayed differential performance increments across time, a 2 (testing condition: 3-month vs. 6-month group) × 2 (time: baseline vs. reevaluation) × 7 (scales: VIQ, PIQ, FSIQ, Verbal Comprehension Index, Perceptual Organization Index, Working Memory Index, Processing Speed Index) mixed factor multivariate analysis of covariance (MANCOVA) was conducted, with time and scales being repeated factors. Age was entered as a covariate. The multivariate model was used to protect against Type I error. In accordance with suggested practice (Harris, 1985), only significant multivariate effects were evaluated in subsequent univariate analyses.

Regarding the MANCOVA, the main effects of time (Hotellings $T^2(1, 46) = 25.23$, $p < .001$) and scales (Hotellings $T^2(6, 41) = 3.62$, $p < .01$) were significant. Additionally, the interaction

1We examined whether WAIS-III administration by two separate evaluators yielded differential test performance across time. We entered IQ and Index Scores into a 2 (same examiner vs. different examiner) × 2 (3-month vs. 6-month reevaluation) × 7 (VIQ, PIQ, FSIQ, VCI, POI, WMI, & PSI) mixed factor ANOVA, with the latter two factors being repeated within participants. No effect associated with examiners was significant. As such, there is no evidence that having different WAIS-III administrators yielded differential performance across time.
of time and scales was significant (Hotellings $T^2$; 41; 41, $p < .01$). No other effect was significant, including the covariate effect of age or the between-groups effect of testing condition. Consequently, follow-up univariate analyses consisted of separate one-way within-subject ANOVAs in which time (baseline vs. reevaluation) served as the repeated factor. ANOVAs were computed for each of the seven WAIS-III scales.

As reported in Table 2, performance on nearly all of the WAIS-III Index Scores improved significantly from baseline to reevaluation. Specifically, the participants showed a mean increase of approximately 3, 11, 6, 4, 8, and 7 points on VIQ, PIQ, FSIQ, VCI, POI, and PSI, respectively. Although scores on the Working Memory Index increased approximately 2 points, this was only marginally significant ($p = .09$).

### Reliable Change Intervals
These analyses indicate that statistically significant improvement occurs on the WAIS-III IQ and Index Scores across 3- and 6-month intervals. Nevertheless, the analyses fail to delineate the magnitude or relevance of the improvement across time. In order to address this issue, reliable change indices were computed (cf. Jacobson & Truax, 1991). These reliable change indices provide the base-rate range of change on a test, and account for random fluctuations across time and the influence of prior testing. These reliable change estimates have proven useful for distinguishing practice effects from genuine recovery of function, and are integral to studying neuropsychological change in clinical samples (cf. Chelune, Naugle, Luders, Sedlak, & Awad, 1993).

As proposed by Charter (1996), reliable change estimates were based on the standard error of prediction (SEP). Whereas the standard error of the difference (Jacobson & Truax, 1991) is commonly used to assess the magnitude of change (cf. Chelune et al., 1993), such a method assumes that the errors between the tests at baseline and readministration are unrelated. Although this assumption is reasonable when assessing change on measures of personality or attitudes, it is untenable with respect to measures of ability. The SEP is computed as follows:

$$SEP = SD_{Y_2}((1 - r_{Y_1Y_2}^2)^{1/2}).$$

$SD_{Y_2}$ is the standard deviation of scores during the second assessment interval, and $r_{Y_1Y_2}$ is the correlation between test scores across the assessment intervals. Two confidence intervals were calculated. A 90% confidence interval was computed by multiplying the SEP by $±1.64$, and then summing the positive and negative values with the estimated true retest score. This interval is consistent with previous investigations (e.g., Chelune et al., 1993), and allows a 5% interval at both tails of the sampling distribution of change scores. Additionally, a more liberal, 50% confidence interval was computed by multiplying the SEP by $±.70$. This interval leaves a 25% interval at both tails of the sampling distribution. This interval

<table>
<thead>
<tr>
<th>WAIS-III IQs</th>
<th>Baseline</th>
<th>Retest</th>
<th>$F$-Value for main effect of reevaluation</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIQ</td>
<td>111.0 (11.5)</td>
<td>114.8 (11.5)</td>
<td>21.15</td>
<td>.30</td>
</tr>
<tr>
<td>PIQ</td>
<td>105.4 (12.5)</td>
<td>116.0 (14.4)</td>
<td>125.25</td>
<td>.72</td>
</tr>
<tr>
<td>FSIQ</td>
<td>109.4 (11.6)</td>
<td>115.04 (12.1)</td>
<td>95.69</td>
<td>.66</td>
</tr>
<tr>
<td>VCI</td>
<td>111.5 (11.9)</td>
<td>115.8 (12.3)</td>
<td>20.28</td>
<td>.29</td>
</tr>
<tr>
<td>POI</td>
<td>106.1 (14.1)</td>
<td>114.4 (14.1)</td>
<td>55.39</td>
<td>.53</td>
</tr>
<tr>
<td>WMI</td>
<td>106.9 (12.4)</td>
<td>108.6 (13.1)</td>
<td>2.93</td>
<td>.06</td>
</tr>
<tr>
<td>PSI</td>
<td>109.3 (13.0)</td>
<td>116.4 (14.5)</td>
<td>32.74</td>
<td>.40</td>
</tr>
</tbody>
</table>

Note: VIQ = Verbal IQ; PIQ = Performance IQ; FSIQ = Full Scale IQ; VCI = Verbal Comprehension Index; POI = Perceptual Organization Index; WMI = Working Memory Index; PSI = Processing Speed Index. Values reflect mean scores at baseline and retest for the entire sample of 51 participants. Standard deviations appear in parentheses. For the $F$-values, $df = (1, 50)$. With the exception of WMI, all $ps < .001$. The WMI $F$-value was not significant.
is reported to assist in decision-making wherein
the risk of making a false positive is less signifi-
cant. Insofar as obtained retest scores exceed the
confidence intervals, genuine change in perfor-
mance is presumed to have occurred. The esti-
mated true score was computed as follows
(Charter, 1996):

\[ Y_{\text{TRUE}} = M + r(Y_{\text{OBS}} - M). \]

In the above equation, \( M \) is the sample mean of
the test, \( Y_{\text{OBS}} \) is the actual score obtained by an
individual, and \( r \) is the reliability coefficient. In
the present study, the test-retest coefficients for
each index were used to estimate true scores upon
reevaluation. Table 3 contains the test-retest coef-
ficients, SEPs and resulting 90% and 50% con-
fidence intervals for each IQ and Index Score
for the entire sample. Because the rate of im-
provement did not differ significantly between the
3- and 6-month reevaluation groups, reliable
change index data are collapsed across these
two groups.

The data in Table 3 demonstrate that a relat-
ively wide range of retest scores may fall within
the 90% confidence intervals. For example, when
utilizing the 90% confidence interval, scores may
range around the estimated true scores by as much
as 9, 11, and 9 points on VIQ, PIQ, or FSIQ,
respectively and reflect error and practice effects
rather than meaningful change. Owing to this,
much variation in test scores may be expected to
occur across 3 and 6 months, and such changes
are due entirely to the effects of practice and
measurement error. Notably, we evaluated
whether any performance during the reevaluation
exceeded the confidence intervals of change. With
the 90% intervals, none did so, but one Working
Memory Index score improved sufficiently that it
exceeded the 50% confidence interval. These
findings suggest that meaningful change typically
failed to occur in this sample of neurologically
normal individuals.

To utilize the confidence interval for an
individual, the estimated true retest score is
computed. The relevant confidence interval is
then applied to the resulting true score. Insofar as
the observed retest score exceeds the upper or
lower limits of the desired confidence interval,
then meaningful change is presumed to have
occurred.

**DISCUSSION**

As stated earlier, research published in the WAIS-
III Technical Manual addresses the issue of prac-
tice effects over a 35-day span. Prior to the current
study, little was known concerning the magnitude
of practice effects on the WAIS-III over longer
time intervals such as 3 or 6 months. The data
reveal that significant improvement occurred
across these time intervals on nearly all of the
WAIS-III IQ scores. With the exception of the
Working Memory Index, significant changes on
the WAIS-III Index Scores were also observed
across time. Furthermore, the magnitude of
change on the IQs and Index Scores was relatively
large, and, as demonstrated by the reliable change
indices, wide variation in performance was com-
mon across time.

Notably, there was no significant interaction of
retest group and time. As such, the average
increases in WAIS-III scores during a 3-month
reeexamination are commensurate with those
observed during a 6-month retest. Related to this issue, it is interesting to note that the increases observed here are similar to those previously reported for the WAIS–R (Kaufman, 1990) and the WAIS-III (Wechsler, 1997). For instance, in the WAIS-III technical manual, the 16–29 age group increased approximately 3, 8, and 6 points across 35 days on the VIQ, PIQ, and FSIQ, respectively (Tulsky & Zhu, 1997). In the current study, participants were approximately the same age, but were reexamined over much longer time intervals. Nonetheless, on the VIQ, PIQ, and FSIQ, they displayed 3-, 11-, and 6-point increases, respectively. As such, there seem to be few differences in WAIS-III score increments across 1-, 3-, or 6-month retest periods.

These findings may appear counter-intuitive, as it initially may seem unlikely that a single exposure to test stimuli would influence subsequent performance, even after periods of 3 or 6 months. Following such long periods, test stimuli may seem to be easily forgotten. However, the data suggest that neurologically normal individuals are able to retain knowledge of testing over extended intervals, thereby influencing subsequent performance.

To account for these increments, it seems likely that individuals recall some aspects of the WAIS-III such as test demands and effective test-taking strategies. Upon reexamination, recollection of these strategies and procedures may have resulted in an overall improvement on test scores. Along with retention of effective strategies, other aspects of the testing situation may have yielded improved performance. For instance, potential anxiety associated with the examination process may have diminished upon reexamination. Hence, examinees may have been more comfortable and better able to focus attention on testing tasks during the second assessment period (cf. Anastasi & Urbina, 1997).

Despite probable retention of procedural knowledge concerning the WAIS-III, performance increments were not uniform across all IQs and Index Scores. Specifically, retesting did not yield improved performance on the Working Memory Index. This suggests that subtests comprising this index are not subject to significant practice effects across 3 or 6 months. In contrast, PIQ scores increased the most of any index. Tasks comprising the Performance scale generally reflect the ability to solve novel problems. Once this novelty is dissipated, responses to the test may become more automatic. For instance, on Object Assembly or Picture Arrangement, individuals may readily recall or recognize the correct solution during reexamination, thereby resulting in quicker and more automatic responses. Accordingly, the current findings imply that subsequent retesting diminishes the novel characteristics of the WAIS-III.

These findings have implications for empirical investigations using the WAIS-III. Because practice effects are apparently common on the WAIS-III, detection of change in clinical status may be obscured. To compensate for this potential difficulty, McCaffrey and Westervelt (1995) suggest using dual baseline assessments. They hypothesize that, upon a second administration, practice effects may plateau, and all subsequent changes in performance will primarily reflect the influence of treatment. Alternatively, inclusion of a no-treatment control group in intervention studies might yield indirect estimates of practice effects. Specifically, any change in performance among a no-treatment control group might provide an estimate of variation due to measurement error and practice effects.

With respect to clinical practice, multiple baselines or no-treatment control groups are impractical or irrelevant. Accordingly, the reliable change index data may hold some utility. Specifically, the estimates of reliable change may provide a benchmark for discriminating patient change from practice effects and measurement error. Yet, it should be acknowledged that the reliable change data might not apply to clinical samples. By definition, practice effects require memory of previous testing. Because clinical patients often are amnesic, they may be less able to remember details of the WAIS-III than the neurologically normal individuals tested in the current study (cf. Groth-Marnat, Gallagher, Hale, & Kaplan, 2000). Thus, clinical samples may show smaller increments in WAIS-III scores across time. Further complicating matters, contributions of practice effects and actual clinical change tend to be confounded in certain clinical
samples (e.g., head injury, cerebral vascular accidents). Because retesting and recovery occur during the same time intervals, the distinction of practice effects from genuine improvement may be difficult to make. Hence, the current data may not present an ideal method for estimating reliable change in clinical samples, but they may provide a reasonable option.

Prior to concluding, it is important to recognize the potential limitations of the present study. Specifically, the participants tended to be predominantly young Caucasians. They were relatively well-educated and possessed mean baseline IQs that were average to high average. It remains uncertain whether the current findings will generalize to other demographic groups such as elderly, below-average IQ, or ethnic minority groups (cf. Horton, 1992; Kaufman, 1990; Lezak, 1995). Typically, practice effects are less robust in older than in younger adults (Kaufman & Lichtenberger, 1999; McCaffrey et al., 1992; McCaffrey & Westervelt, 1995). As such, older individuals may not show WAIS-III increments as large as those observed in the present study. In support of this hypothesis, Tulsky and Zhu (1997) found that practice effects across 1 month tended to be smaller in older than younger age groups (cf. Horton, 1992; Kaufman, 1990; Lezak, 1995). Typically, practice effects are less robust in older than in younger adults (Kaufman & Lichtenberger, 1999; McCaffrey et al., 1992; McCaffrey & Westervelt, 1995). As such, older individuals may not show WAIS-III increments as large as those observed in the present study. In support of this hypothesis, Tulsky and Zhu (1997) found that practice effects across 1 month tended to be smaller in older than younger age groups. Yet, despite these potential limitations, the current findings provide potentially useful information to clinicians and researchers who wish to employ the WAIS-III in repeated examinations across 3- or 6-month intervals. Notably, these findings indicate which WAIS-III indices are susceptible to practice effects, and they provide compelling evidence that even apparently substantial changes in performance may reflect the confounding effects of practice and measurement error.

REFERENCES