

Frequently Asked Questions

Regression Testing

I administered the WISC-IV to a student who scored 65 on each of the VCI, PRI, WMI, and PSI indexes, but his FSIO was 57. Shouldn't it be 65?

Many people find this result counterintuitive, but it is correct. First, consider that the FSIQ is used to predict the student's true intelligence and does not correlate perfectly with it. Then consider that the index scores are composed of fewer subtests than the full scale IQ score and do not correlate perfectly with the FSIQ. In this case, if the student's true IQ is 57, then his or her index scores should be higher than 57 due to the effect of regression toward the mean. On the other end of the continuum, the opposite is true. If a student's FSIQ is 147, there is a greater probability that his or her index scores will be lower than the FSIQ.

This effect can be found in the composite score norms tables of many tests of cognitive ability, though the strength of the effect depends on several factors, including the number of subtests entering the composite, and the distance of the subtest scores from the mean, and the correlation among those subtests.

When a composite is made up of more subtests, the effect is larger. It is rarer to score about 2 standard deviations below the mean on each of the 10 subtests that compose the FSIQ than on each of the two or three subtests that compose an index score. This is why the effect is more pronounced for the FSIQ than for any of the four Indexes.

The further a score is from the mean, the larger the effect. This is because it is rarer to score about 2 standard deviations from the mean on all 10 core subtests than it is to score 1 standard deviation from the mean all 10 subtests. The effect is usually more pronounced at 2 standard deviations from the mean than at 1 standard deviation from the mean. In WISC-IV, the effect is largest at approximately 2 standard deviations above or below the mean. Beyond this point, the minimum and maximum possible scores constrain the effect.

The regression toward the mean effect is stronger when there is a lower correlation among the subtests that make up the composite score. That is why the effect is stronger in WISC–IV than it was in WISC–III. Though the inter-correlations among the 10 core subtests that constitute the WISC–IV FSIQ are high, they are not as high as those for the WISC–III subtests.

These slightly lower correlations are related to the greater diversity of construct coverage among the core subtests in WISC–IV. This expanded construct coverage improves clinical utility, but weakens the intercorrelations among the core subtests, which in turn, increases the effect of regression toward the mean.

The mean scores for the MR groups reported in the WISC-IV Technical Manual are FSIQ = 60 (Mild MR group) and FSIQ = 46 (Moderate MR group), respectively. These two clinical studies provide evidence in support of the psychometric integrity of the normative data in this range of the distribution.

In a separate study of 84 children selected from the combined MR sample groups, the WISC–IV Index and FSIQ scores were compared with previously obtained WISC–III scores (mean re-test interval = 6 months). As shown in the following table, the results of the WISC–IV are consistent with the WISC–III scores. The mean FSIQ scores are almost identical. The corrected correlations between the composite scores of the two versions range from .83 to .90. In addition, 68% and 81% of the WISC–IV FSIQ scores are within the 95% and 99% confidence interval of the WISC–III FSIQ scores, respectively. Such results are similar to the test-retest results reported in the WISC–III and WISC–IV manuals, and supply an additional line of evidence in support of the validity of the test scores in this portion of the IQ range.

Composite Score Consistency Between the WISC-IV and WISC-III Based on a Sample of Students with Mental Retardation								
	WISC-IV			WISC-III				
	n	Mean	SD	n	Mean	SD	Discrepancy	d
VIQ/VCI	78	61.8	11.4	77	58.9	8.9	2.9	.019
PIQ/PRI	84	61.3	12.2	76	60.1	9.4	1.3	0.08
FSIQ	72	56.0	11.3	84	56.0	8.6	0.0	0.00

 r^c = corrected correlation

Test Framework and Revisions

How is the FSIQ on WISC-IV different than the FSIQ on WISC-III?

Compared to the WISC–III, the WISC–IV FSIQ deemphasizes crystallized knowledge (Information is supplemental), and increases the contribution of fluid reasoning (Matrix Reasoning and Picture Concepts), working memory (Letter–Number Sequencing), and Processing Speed (both Coding and Symbol Search). The WISC–IV FSIQ is comprised of all 10 subtests that comprise the four index scores, including additional measures of working memory and processing speed. The WISC–III FSIQ included only one measure of processing speed and one measure of working memory in the FSIQ.

The new indices are made up either of two or three subtests. Are these as reliable as the former VIQ and PIQ?

Yes. Although the indices contain fewer subtests than in WISC–III VIQ and PIQ, the reliability is just as high for the indices as it was for the VIQ and PIQ. This is mostly due to the removal of subtests that demonstrated relatively lower reliability than those subtests that were retained and the highly reliable subtests that were added.

Picture Completion loads higher on the PRI than Picture Concepts. Why was Picture Concepts chosen as the core subtest?

Factor loadings were one of the criteria considered when deciding which subtests are core and which are supplemental; equal emphasis was placed on the clinical utility and the breadth of construct coverage than on factor loading. Picture Completion and other, more traditional measures of perceptual ability, measure visual discrimination and attention to visual detail, which is a lower order cognitive ability than fluid reasoning.

Cancellation has a high factor loading on PSI. Why was Cancellation not included as a core subtest?

Factor loadings are one of the criteria considered when deciding which subtests are core and which are supplemental. Although Cancellation shows preliminary evidence of clinical utility in terms of the difference in performance between the structured and random presentations, it correlates lower with FSIQ than the subtests that were included as core subtests, and would involve the purchase of an additional response booklet for every administration.

Is Digits Backward a better Working Memory task than Digits Forward?

Digit Span Forward and Digit Span Backward tap distinct but highly interdependent neurocognitive functions. Digit Span Forward primarily taps short-term auditory memory while Digit Span Backward measures the child's ability to manipulate verbal information while in temporary storage. Digit Span Forward is a precursor ability to Digit Span Backward for normally developing children. The separate scores allow practitioners to evaluate possible dissociation of these functions in disordered populations.

Can I substitute the supplemental subtests for a core subtest?

Yes, you can substitute one supplemental subtest per index. However, you can only substitute a maximum of two subtests total to retain the validity of the FSIQ.

With five supplemental subtests available, can I give all the supplemental subtests and use the highest scores?

No. The supplemental subtests are designed to be used as substitutes in computing composite scores for spoiled or invalidated subtests. If they are given in addition to the core subtests they can provide additional information on cognitive functioning; however, decisions should not be made after the tests are given as to which subtests will be used to derive composite scores. If a supplemental subtest is used in place of a core subtest for clinical reasons, this decision should be made prior to administration of the WISC–IV, not after the scores are derived. For example, an examiner working with a motorically-challenged child may decide prior to testing that she or he will administer Picture Completion as a substitute for Block Design. Supplemental subtests are also useful when the scores within an index are widely discrepant. The additional information from the supplemental subtest can help tease out factors contributing to disparate results.

Was the WISC-IV designed to follow the CHC factors?

The development of the WISC–IV was significantly influenced by current research into neurocognitive information processing models, and the creation of new subtests was equally guided by clinical research and factorial data. The Wechsler four-factor structure was first introduced as an option within the WISC–III (1991) and subsequently included in the WAIS–III (1997). The WISC–IV (2003) strengthens the Wechsler four-factor model, and removes its status as optional. Nonetheless, the WISC–IV subtests measure constructs that could be described using common CHC terms such as fluid reasoning (MR, PCn, SI, WR), quantitative knowledge (AR), crystallized knowledge (IN, VC), short term memory (DS, LN, AR), visual processing (PC, BD), long-term storage and retrieval (IN, VC), and processing speed (CD, SS, CA).

Why was Picture Arrangement dropped?

Picture Arrangement was dropped for a variety of reasons. Efforts were made to decrease the emphasis on time bonuses where possible. Picture Arrangement scores were heavily dependent on time. Dropping Picture Arrangement also reduced the time required to administer a WISC-IV. Finally, in terms of user friendliness, dropping Picture Arrangement meant removing a subtest with multiple pieces that could be lost or administered inconsistently. Ultimately, in order to make room for new subtests, some difficult choices needed to be made.

Why was Object Assembly dropped?

Object Assembly was also dropped for a variety of reasons. There was an emphasis on decreasing dependence on time bonuses. The removal of Object Assembly was also aimed at reducing the amount of time it takes to administer the WISC-IV. Dropping Object Assembly allowed for the PRI to be a much purer reasoning index. Finally, in terms of user friendliness, dropping Object Assembly meant removing a subtest with multiple pieces that could be lost or administered inconsistently. As was the case with Picture Arrangement, some difficult choices needed to be made in order to make room for the new subtests.

Why was Arithmetic dropped as a Core subtest?

Arithmetic was dropped because of the significant math load that confounded the meaning of the result. However, it was retained as a Supplemental subtest because it is an excellent measure of Working Memory. Working Memory must be assessed with some content load. For children with grade appropriate math skills, Arithmetic is an excellent measure of Working Memory. However, results are difficult to interpret for children weak in math skills. Mental arithmetic is a task with considerable ecological validity because it is something many people are frequently called upon to do in real life.

General Administration and Scoring

Which GAI tables should we use? Four different GAI tables are provided in various sources, and they are not the same. Which one is endorsed by PsychCorp? The four sources are: WISC-IV Clinical Use and Interpretation (Prifitera, Saklofske & Weiss, 2005), Essentials of WISC-IV Assessment (Kaufman & Flanagan, 2004); WISC-IV & WPPSI-III Supplement Sattler & Dumont, 2004), and on the Dumont & Willis website.

The GAI tables provided by Prifitera, Saklofske & Weiss (2005) are the only GAI tables supported by The Psychological Corporation. That is because they were created using the actual WISC-IV standardization sample (n = 2,200), whereas the GAI tables provided in other sources were created using a statistical approximation (Tellegen & Briggs, 1967). Further, there are differences among the various tables based the Tellegen & Briggs formula depending on whether the scores were derived from sums of scaled scores on the subtests, or sums of standard scores on the Indexes, and all of those tables will differ from tables derived directly form the original normative data. The Tellegen & Briggs formula is appropriate for use in situations were the actual data are not available, and the tables provided by others were generated while practitioners were waiting for the actual tables to be created. Now that Prifitera et. al., have provided the actual tables, any GAI tables found in other sources should be considered approximations. The Tellegen & Briggs formula underestimates scores in the upper portion of the distribution and overestimates scores in the lower portion of the distribution. On average, this difference is about 2 to 3 points, but can be as much as 6 points in individual MR and gifted cases. Thus, practitioners are advised to use the GAI formulas found in WISC-IV Clinical Use & Interpretation (Prifitera, Saklofske, & Weiss, 2005). The book is published by Elsevier Science and now available at the www.elsevier.com, or by selecting Pearson Customer Service from the dropdown of our Contact Us page.

Why are some 0 point or 1 point responses on the verbal subtests not queried?

In standardization it was determined that querying certain responses did not result in any additional information. If you feel the child has more knowledge, based on your clinical judgment, the child's performance on surrounding items, and other observations during the administration, you have the option to query. However, clearly wrong responses should not be queried. In addition, the responses marked with a Query in the manual, must be queried.

Why are there separate norms for Block Design with and without time bonuses?

Practitioners have suspected that some children who emphasize accuracy over speed of performance may score lower on Block Design because of time bonuses, while others believe that faster performance reflects a higher level of the ability being measured. The separate scores allow practitioners to evaluate these hypotheses with individual children. Practitioners should be aware that most children in the standardization sample achieve very similar scores on Block Design and Block Design Non-Time Bonus. In fact, a difference as small as two points is considered rare.

What is the rule of thumb for clinical significance in base rates?

In general, you should use the rule of 10%. Once you get a base rate that is less than 10%, you should begin to do additional hypothesis testing to confirm or disprove your conclusions. However, if there are medical reasons to expect certain discrepancies, such as a previous traumatic brain injury, then even 15% or higher could be meaningful.

What scores do I use if I want to do a discrepancy analysis?

The VCI is the functional equivalent of the VIQ. Similarly, the PRI is the functional equivalent of the PIQ. You should use the VCI and PRI as you would use the VIQ and PIQ. One significant improvement in the WISC-IV is the ability to do a number of other discrepancy analyses. For example you could look at VCI versus WMI or PSI. You could compare PRI with WMI or PSI. Working Memory can be compared directly with PSI. You can also compare Block Design performances with and without time bonuses, Digits Forward versus Digits Backward, and Cancellation's unstructured trial with the structured trial.

How do you interpret the difference between Picture Concepts and Similarities?

Although both Picture Concepts and Similarities are measures of conceptualization and categorization, the stimulus and response modalities are different. Similarities uses a verbal stimulus and requires a verbal response. Picture Concepts uses a visual stimulus and requires either a motor or a verbal response. Although performance of both tasks may include internal verbal mediation, Picture Concepts does not require that the examinee verbally express his or her categorical concept. Differences help identify children who have good conceptualization skills but are less adept at articulating their rationale.

Clinical and Special Group Performance

Why is reliability lower for gifted children and children with mental retardation than children in the standardization sample?

It is a consistent finding that the restriction in the range of scores these groups obtain frequently results in lower reliabilities.

Are there profiles typical of clinical disorders?

In general, the answer is no. However, ongoing research may identify certain characteristics of cognitive functioning for specific clinical disorders. While specific profiles are not diagnostic of particular disorders, working memory and processing speed are implicated in a variety of psychoeducational and neuropsychological disorders.

Do children with learning disabilities score lower on WMI and PSI?

Studies reported in the Manual suggest that children with learning and/or attention disorders tend to perform lower on tasks that measure working memory and processing speed.

I retested a gifted student using WISC-IV and the scores were lower than previously reported on WISC-III. Why is this?

This is due to the difference in the core subtests between WISC-III and WISC-IV; core subtests in the WISC-IV reflect the increased emphasis on fluid reasoning, working memory, and processing speed in more recent

conceptualizations of intelligence. The removal of the Information subtest from the core battery reduces the contribution of crystallized knowledge to the FSIQ. The addition of Picture Concepts and Matrix Reasoning subtests results in a much stronger element of fluid reasoning. On the WMI (formerly called FDI on WISC–III), Letter-Number Sequencing replaces Arithmetic, a subtest on which intellectually gifted students tended to score highly, due to school based learning of mathematical skills. One additional processing speed subtest was added to the core battery (SS). Gifted students tend not to score as high on processing speed subtests relative to other indices, perhaps due to an approach to problem solving that stresses accuracy over speed of performance. In addition to the difference in the core subtests on WISC-III and WISC-IV, the norms for the newer test are slightly harder due to the Flynn effect. Although, some children exhibit scores that regress to the mean upon retesting, analyses of the standardization data from the WISC-III and WISC-IV indicate that the same percentage of children, approximately 2%, is identified as gifted based on the FSIQ. However, the same children may not be identified due to the shift in the conceptualization of intelligence reflected in the core subtests that contribute to the WISC-IV FSIQ.

To meet the cognitive requirements for a diagnosis of mental retardation my state requires that the VIQ, PIQ, and FSIQ scores all be below 70 points. How do I do this with WISC-IV which no longer has scores labeled as VIQ and PIQ?

States and other regulatory bodies may update their terminology in the near future. In the meantime, there is a statement on page 4 of the WISC-IV Administration and Scoring Manual that was designed to address this situation. The statement reads as follows: "The terms VCI and PRI should be substituted for the terms VIQ and PIQ in clinical decision-making and other situations where VIQ and PIQ were previously required."

I work in a school district where many different languages are spoken. What do I do with a child who has recently immigrated to the United States and needs to be assessed in a language other than English?

The WISC–III has been adapted and standardized in 16 different countries. For children whose families have recently immigrated, these are the most current, valid tests available in their first language. Versions of these tests can be obtained by contacting PsychCorp and include adaptations for Canada, United Kingdom, France and French-speaking Belgium, The Netherlands and Flemish-speaking Belgium, Germany, Austria and Switzerland, Sweden, Lithuania, Slovenia, Greece, Japan, South Korea, and Taiwan. Use of these adaptations requires an examiner or experienced professional who is fluent in the child's language.

In addition to these, the WISC–IV Spanish Edition is currently in development for use in the United States. Bilingual examiners (Spanish/English) and are urged to call PsychCorp at 1-800-228-0752, ext. 5218 if they interested in participating as examiners in the WISC–IV Spanish Edition standardization study. Standardization projects are underway for English language versions in Australia, England, and Canada that offer local normative information; a French Canadian version is also under development for use in French-speaking Canada.

Subtest Administration and Scoring

Why does WISC-IV start with Block Design?

Although Picture Completion has traditionally been the first subtest administered, it is not a core subtest in the WISC-IV. Block Design was chosen as the first subtest because it is an engaging task that allows the examiner additional opportunity to establish rapport. This is consistent with a recent revision of another Wechsler product, the WPPSI-III, where Block Design as the initial subtest has been well-received by examiners. When testing motorically challenged children, examiners may decide to begin with a different subtest in the interest of rapport.

If you wanted to reduce the effects of speeded performance, why not eliminate time bonus altogether from Block Design?

In general, higher ability children tend to perform the task faster. Without time bonuses, Block Design is not as good of a measure of high ability.

Why is Digit Span placed so early in the subtest order?

In order to avoid interference effects between Digit Span and Letter-Number Sequencing, these subtests were widely separated in the order of administration.

On Picture Concepts, why do some children seem to lose track when three rows are first introduced?

Typically, when children lose the instructional set when three rows are introduced, they have reached the upper limit of their ability on this subtest; they lose track of the instructions and are drawn to the distracters included in each row of items. Children should be prompted as instructed each time this loss of set occurs.

Why do there seem to be multiple responses for some of the items on Picture Concepts?

The Picture Concepts subtest is scored with either 0 or 1 point. The keyed response represents the best single response in terms of the level of reasoning involved.

For example, on more difficult items credit is not given for categories involving color or shape, emphasis is placed on underlying function. The keyed response was determined through years of research in Pilot, Tryout, and Standardization phases of development. The categories children provided, the ability level of children choosing specific responses, and relationships to performance on Similarities and Matrix Reasoning were all used to determine the keyed response.

Why have picture items been added to the Vocabulary subtest?

These items were added to increase the floor and provide a way to assess very low functioning children.

In the Letter-Number Sequencing subtest, the examinee is instructed to give the numbers in order first and then the letters in order. Why is credit awarded if the examinee gives the letters first in order and then the numbers in order?

There is a distinction between reordering and sequencing: Reordering involves placing the numbers as a group

prior to the letters as a group, and sequencing involves placing the numbers in numerical order and the letters in alphabetical order—regardless of which grouping comes first. This distinction is reflected in the prompt given and relates directly to how a trial is scored. If a child states the letter first on Item 1, the child is prompted to reorder the group, however, despite the prompt, the child still receives credit for his or her original answer because the response is one of the two correct responses listed. Items 4 and 5 prompt the child to place the numbers or letters in sequential order; on these items no credit is awarded if the child has to be prompted because, unlike Item 1, the original sequence is not one of the correct responses listed for these items. You may prompt the child once for Items 1, 4, and 5; you cannot prompt a child on any of the other items for this subtest. Regardless of how the child reorders the numbers and letters, he or she is using working memory in order to place the numbers in sequence and the letters in sequence. Data analyses of the standardization sample showed that the task is equally difficult when either numbers or letters are given first. The reason for instructing examinees to give the numbers first is to provide them with a set or structured way of approaching the task, which is especially helpful for young children or children who have difficulty structuring their own work. This is the same scoring method used for Letter-Number Sequencing on WAIS–III.

On Letter-Number Sequencing, how do I score a child's response after a prompt is given on Items 1, 4, and 5?

As noted in the third bullet under General Directions on page 126 of the Administration and Scoring Manual, certain responses to specific trials on Items 1, 4, and 5 require a prompt to remind the child of what the task is for this subtest. The prompt for Trial 1 of Item 1 is designed to remind the child to say the numbers first and then the letters. If the child forgets to say the number first (i.e., repeats exactly what you say), award credit for the trial as indicated and provide the prompt. Because the child received credit for his or her initial response to this trial, it is not necessary to award additional credit if the child attempts to correct his or her initial response after the prompt.

Trial 2 of Item 4 is the first trial in which the child is required to alphabetically sequence the letters to produce a correct response. If the child provides either of the specified incorrect responses by forgetting to alphabetically sequence the letters, provide the prompt as indicated. If the child provides a correct response to the trial after the prompt, do not award credit for the trial.

Similarly, Trial 1 of Item 5 is the first trial in which the child is required to sequence the numbers to produce a correct response. If the child forgets to sequence the numbers and provides either of the designated incorrect responses, provide the prompt as indicated. If the child provides a correct response to the trial after the prompt, do not award credit for the trial.

On the Letter-Number Sequencing subtest, a child can simply mimic the examiner and earn credit on the first 10 trials. Is this really working memory?

The early items measure short-term auditory memory, which is a precursor skill to working memory. The 6–7 year old norms demonstrate that children scoring 10 raw score points obtain above average scaled scores; this reflects the developmentally appropriate use of short-term memory prior to the exhibition of working memory. Thus, for younger children, Letter-Number Sequencing may assess short-term memory, a prerequisite skill for the development of working memory. The item set and norms reflect this change as children develop working

memory. This is analogous to the difference between Digit Span Forward and Backward, which assesses short-term memory and working memory, respectively. Performance on the early items of Letter-Number Sequencing in younger children may be related to performance on Digit Span Forward with any differences potentially attributable to automaticity of letters as compared to numbers.

The answer on Matrix Reasoning item #26 does not appear to be the only possible answer. Why wasn't "2" given credit?

Item #26 is the second 3 X 3 item. On the first 3 X 3 item (#24), children learn to apply the same transformation from cell #1 across to cell #2, and again from cell #2 to cell #3. If the child follows the pattern from #24, they answer correctly (1) on item #26. Children can arrive at a different answer (2) if they use one transformation rule from cell #1 to cell #2, and a different one from cell #2 to cell #3. This is not the most parsimonious solution, and analyses indicated that children who arrived at the correct response (1) had higher ability levels.

What should I do if a child writes too faintly to be seen through the Cancellation scoring template?

You don't need the scoring template to score the subtest. If necessary, remove the template and simply count each animal with a mark through it and each non-animal with a mark through it. You should make sure to double-check your work.

Is color-blindness a factor in performance on the Cancellation subtest?

No. Although the Cancellation task utilizes color as a visual distracter, making it possible that children who are color blind will be less distracted by the bright colors or have greater difficulty differentiating objects of various colors, it is recognition of the shapes of the objects that is required to place them in categories.

What does it mean if a child guesses right on the first clue of Word Reasoning?

Children are more likely to guess correctly on the easier items such as those that appear in the first half of the item set, especially Item 9. The more difficult items found in the second half of the item set show a very low percentage of correct responses to the first clue. In order to respond correctly, even on the first clue, the child must use deductive reasoning; that is, on the first clue the child has to narrow the potential responses to those that fit a search set defined by the clue and then make a reasoned guess from the range of responses within the set.

It is possible that a child who consistently guesses correctly on the first clue may have taken the test recently, or may have been coached on the correct responses.

Some of the responses I get on Word Reasoning seem correct to me, but are listed as incorrect. What was the rational for determining correct and incorrect responses?

Some of these responses may have been given 1 point in a 0, 1, 2 point scoring rubric. Such responses may be correct, or partially correct, but do not represent a high level of abstract reasoning. They also tended to be given

by children with lower ability. Not all possible responses are included in the examples, however, and the examiner may give credit for a response not listed if she or he determines that it is at the same level of abstraction as the credited responses.

Related Instruments

What will it take to upgrade to the integrated?

For an introductory period, the WISC-IV price plus the upgrade price will equal the price of the complete WISC-IV Integrated.

Will the WISC-IV Spanish be a "true" Spanish version or just a translation?

The WISC-IV Spanish will be a "true" Spanish version, developed from the "ground up" in Spanish. Some items, where appropriate, will be adopted from WISC-IV in English. However, the majority of the instrument will be developed in Spanish. A panel of scholars is working together to minimize the dialectic and regional differences from the variety of Hispanic nationalities that will be included in the standardization sample. Finally, the resulting scores will be equated to the WISC-IV norms so that Hispanic children can be measured and compared to the same scale.



