



LIGHTING THE PATHWAYS  
TO LEARNING

**Research Report**  
**No. 2013-2**

# Recent Trends of Test- Taker Characteristics & Mean Scores on the ERB Adapted WPPSI-III Assessment



# Recent Trends of Test-Taker Characteristics & Mean Scores on the ERB Adapted WPPSI-III Assessment

Tammy J. Trierweiler, Ph.D.



## Contents

Introduction.....	5
Statement of the Problem.....	5
Purpose of the Current Study.....	6
Methods.....	7
Results.....	9
Summary .....	33
Limitations & Future Directions.....	36
References.....	38
Appendix A.....	39
Appendix B .....	41
Appendix C .....	45
Appendix D.....	49



## Tables

Table 1. <i>Mean Age in Years &amp; Distribution of Gender by Administration</i> .....	10
Table 2. <i>Percent of WPPSI-III Test Takers by Grade at Time of Testing</i> .....	11
Table 3. <i>Mean Age in Years of WPPSI-III Test Takers by Grade at Time of Testing</i> .....	11
Table 4. <i>Mean WPPSI-III Composite Scores across Administrations</i> .....	16
Table 5. <i>Mean Gain Scores &amp; Standardized Mean Differences</i> .....	18
Table 6. <i>Percentage of children scoring at or above the 90<sup>th</sup> and 98<sup>th</sup> percentiles – Full Scale</i> ..	20
Table 7. <i>Percentage of children scoring at or above the 90<sup>th</sup> and 98<sup>th</sup> percentiles – V &amp; P</i> .....	22
Table 8. <i>Difference test of Proportions across Administrations – Full Scale</i> .....	26
Table 9. <i>Difference test of Proportions across Administrations – Verbal</i> .....	26
Table 10. <i>Difference test of Proportions across Administrations – Performance</i> .....	27

## Figures

Figure 1. <i>Mean Full Scale Scores by Gender</i> .....	13
Figure 2. <i>Mean Verbal Scores by Gender</i> .....	13
Figure 3. <i>Mean Performance Scores by Gender</i> .....	13
Figure 4. <i>Mean Full Scale Scores by Grade at Testing</i> .....	14
Figure 5. <i>Mean Verbal Scores by Grade at Testing</i> .....	14
Figure 6. <i>Mean Performance Scores by Grade at Testing</i> .....	14
Figure 7. <i>Mean Full Scale Scores across Administrations</i> .....	17
Figure 8. <i>Mean Verbal &amp; Performance Scores across Administrations</i> .....	17
Figure 9. <i>Percentage of Children at or above the 90<sup>th</sup> Percentile – Full Scale</i> .....	20
Figure 10. <i>Percentage of Children at or above the 90<sup>th</sup> Percentile – Full Scale</i> .....	21
Figure 11. <i>Percentage of Children at or above the 90<sup>th</sup> Percentile – Verbal</i> .....	23
Figure 12. <i>Percentage of Children at or above the 98<sup>th</sup> Percentile – Verbal</i> .....	23
Figure 13. <i>Percentage of Children at or above the 90<sup>th</sup> Percentile – Performance</i> .....	24
Figure 14. <i>Percentage of Children at or above the 98<sup>th</sup> Percentile – Performance</i> .....	24
Figure 15. <i>Percentage of Children at or above the 90<sup>th</sup> Empirical Percentile – Full Scale</i> .....	29
Figure 16. <i>Percentage of Children at or above the 98<sup>th</sup> Empirical Percentile – Full Scale</i> .....	29
Figure 17. <i>Percentage of Children at or above the 90<sup>th</sup> Empirical Percentile – Verbal</i> .....	30
Figure 18. <i>Percentage of Children at or above the 98<sup>th</sup> Empirical Percentile – Verbal</i> .....	31
Figure 19. <i>Percentage of Children at or above the 90<sup>th</sup> Empirical Percentile – Performance</i> .....	31
Figure 20. <i>Percentage of Children at or above the 98<sup>th</sup> Empirical Percentile – Performance</i> .....	32



## **Introduction**

ERB is a not-for-profit educational resources membership association of independent, public, faith-based, and boarding schools providing admission, achievement and support services for Pre-K through Grade 12.

In New York City, ERB has served as the testing agency for the Independent Schools Admissions Association of Greater New York (ISAAGNY) since 1966. Using tests from its Early Childhood Admission Assessment (ECAA) program — a program that offers assessments for children applying to Pre-K through Grade 4 — ERB meets the testing needs of applicants to ISAAGNY member schools by eliminating repetitive admission testing thereby minimizing the strain on both children and parents.

For those children applying to grades Pre-K through 1, the ERB adapted Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III)<sup>1</sup> is used. This ERB adapted WPPSI-III was created specifically for use as an admission tool and it is systematically different than the nationally normed WPPSI-III that was created for clinical use. The ERB adapted version of the assessment features eight unique subtests, four of which explore verbal skills and four of which explore non-verbal, or performance, abilities. Subtest scores can be used independently, or they may be combined to yield a full scale composite score, which is simply the sum of the child's attained scaled scores in both the verbal and non-verbal domains. The resulting scores from the ERB-modified WPPSI-III assessment are not to be used as to gauge a child's general intelligence, but are to be used in conjunction with school records and personal interviews as an additional tool to help guide admission decisions.

## **Statement of the Problem**

There have been increasing concerns from some schools that participate in the ECAA One to One program that use the ERB adapted WPPSI-III assessment in their admission process that resulting test scores are inflated and that the percentile rankings of the children taking the test have increased dramatically over the recent years. This general perception revolving around

---

<sup>1</sup> Beginning in the 2013-14 administration year, the WPPSI-III has been replaced with the fourth version of the Wechsler Preschool and Primary Scale of Intelligence <sup>TM</sup> (WPPSI-IV).



score increases has brought into question whether or not the use of the ERB adapted WPPSI-III scores is appropriate for use in the admission process.

One driving force behind worries about score inflation on the WPPSI-III is the increased availability of test preparatory materials to families going through the admission process. Many believe that, with the recent and highly publicized accessibility of test preparation materials, these tools are ceding unfair advantages to children who have access to and use these services. In addition, with general public perception about the dangers associated with standardized testing at early ages — combined with surface evaluations of individual school data indicating increased number of children scoring at the upper percentile ranks on the WPPSI-III — a heated debate has commenced about the validity of the use of these scores in the admission process, and many argue that scores may not be proper indicators of actual child ability.

Swaying personal ideologies and public perceptions circling the use of standardized testing in early education, however, is not the aim of this research report. The primary goal is to empirically evaluate recent trends in test-taker characteristics as well as mean WPPSI-III composite scores over a period of nine years. It is assumed that, if accessibility to test preparation materials has skyrocketed in recent years, then specific trends in the data will become apparent. More importantly, no research has evaluated demographic or score trends on the ERB adapted WPPSI-III over an aggregate sample. Most evaluation to date has been done on a school by school basis, not taking into account fluctuations in test-taker characteristics or multiple years of data.

## **Purpose of the Current Study**

Following, the purpose of this study is to evaluate recent trends in test-taker characteristics and mean ERB adapted WPPSI-III scores over nine recent administration periods. This research report will attempt to answer the following questions:

- ❖ **Has the demographic composition of children taking the ERB adapted WPPSI-III changed across years? Total sample size, gender, age, and grade at time of testing are reported.**
- ❖ **Across years, is there evidence of trends in WPPSI-III mean scaled scores by gender or a child's grade at testing?**

- ❖ **Across years, is there evidence of trends in WPPSI-III mean scaled scores across the aggregate sample?**
- ❖ **Have the WPPSI-III mean scores increased substantially over administrations?**
- ❖ **Across years, is there evidence of an upward trend in the proportion of children scoring at the 90<sup>th</sup> and 98<sup>th</sup> percentiles on the WPPSI-III?**
- ❖ **Does the proportion of children scoring at the 90<sup>th</sup> and 98<sup>th</sup> percentiles on the WPPSI-III change significantly over time?**
- ❖ **Across years, is there evidence of an upward trend in the proportion of children scoring at the 90<sup>th</sup> and 98<sup>th</sup> percentiles on the WPPSI-III based on the empirical ERB sample data?**

## **Methods**

The WPPSI™-III is a nationally normed intelligence test used for children from 2 years 6 months to 7 years 3 months of age. As the WPPSI™-III was developed to be a nationally normed intelligence test, not an admission test *per se*, ERB contracted with Pearson Education to develop an adapted WPPSI-III specifically for ERB ECAA purposes (ERB, 2002/2013).

This ERB adapted WPPSI-III includes eight subtests and yields three composite scores: a verbal score, a performance score, and a full scale score. The verbal score reflects skills in understanding verbal information, whereas the performance score reflects skills in solving nonverbal problems. The full scale score is derived from the combination of both the verbal and performance scores and reflects overall thinking and reasoning skills. It is important to highlight here that the scores on the ERB adapted WPPSI-III are *not equivalent* to the nationally reported WPPSI™-III scores used for neuro-cognitive diagnostic assessment. The ERB adapted WPPSI- III scores have been developed specifically for admission testing and should be used with other criterion in the admission process.

ERB adapted WPPSI-III data used in this report was collected over a nine-year administration period (2004-5 to 2012-13). As this study's central goal was to identify characteristics and mean



scores of children applying to Kindergarten, only children applying to Kindergarten were evaluated. Demographic data evaluated across administrations include age, gender, and the grade of the child at time of testing. Children who were missing any composite score data or who had full scale composite scores that fell outside the reported range (e.g., 40 – 160) were also excluded from analyses.

To evaluate trends in test-taker characteristics and WPPSI-III mean score across time, descriptive statistics are presented along with graphical illustrations for each criterion variable of interest. To assess if trends in mean WPPSI-III composite scores are of practical significance, gain scores as well as the magnitude of these gain scores across consecutive administrations are presented.

To assess the magnitude of the change using the gain scores across administrations, standardized mean differences are used as given the large sample sizes across administration, *t*-tests or other general linear models (i.e., ANOVAs and regressions), significance between even the smallest of differences are usually found as sample size is taken into consideration in the calculations.

The standardized mean difference or “effect size” measure (e.g., Cohen’s *d*) involves dividing the difference between two group means (i.e., Time 1 and Time 2) by their average (pooled) standard deviation:

$$\text{Effect Size} = \frac{\text{Mean Scale Score}_{\text{Time1}} - \text{Mean Scale Score}_{\text{Time2}}}{\left( \frac{\text{SD Mean Scale Score}_{\text{Time1}} + \text{SD Mean Scale Score}_{\text{Time2}}}{2} \right)}$$

The interpretation of the effect size can be evaluated using Cohen’s (1988) suggestion that effect sizes of about 0.20, 0.50, and 0.80 are to be considered small, medium, and large, respectively. Generally, effect sizes less than 0.20 are considered trivial, and effect sizes larger than 0.33 are considered to be large enough in magnitude to warrant practical significance.

In addition to evaluating mean trends, the chi-square ( $\chi^2$ ) test is also used to evaluate if the proportions of children scoring in the 90<sup>th</sup> and 98<sup>th</sup> percentiles is significantly different across each consecutive administration.

As the percentile ranks assigned to children taking in the ERB adapted WPPSI-III are based on a nationally normed standardization sample who took the WPPSI-III for neuro-cognitive





diagnoses, the ability to make comparisons between the ERB sample and the standardized sample using the same percentile rankings may be unsuitable.

To circumvent this concern, empirical percentile ranks of the ERB samples across administrations are calculated and evaluated using the observed sample data the following formula (Ley, 1972):

$$\text{Percentile Rank} = \left[ \frac{m + .5k}{N} \right] * 100$$

where  $m$  is the number of children in the sample scoring below a given score,  $k$  is the number obtaining the given score, and  $N$  is the overall sample size.

Empirical percentile ranks are simply comparing children within each administration to their peers within that administration, and assigning percentile ranks based on the observed data—what is sometimes called a ‘local’ norm.

## Results

---

### Has the Demographic Composition of Children Taking the ERB Adapted WPPSI-III Changed Across Years?

---

This section evaluates demographic characteristics of the children taking the ERB adapted WPPSI-III during the Kindergarten admission process. Results illustrate trends for test takers disaggregated by gender as well as the specific grade level these children tested in during their application process. Knowledge of similarities or differences among children from various subgroups is provided.

#### *Age & Gender*

Table 1 indicates the mean age of the Kindergarten admission samples across administrations, as well as the distribution of children by gender. As can be seen, the mean age of children across the nine administrations is relatively consistent across time, although there does appear to be a slight upward trend in the mean age of the children taking the test across the 2009-2010 and the 2012-

2013 years. Children appear to be approximately one to two months older in the most recent administrations than earlier cohorts.

The number of males and females taking the assessment for Kindergarten admission has remained relatively stable across years, with only small variations.

Table 1. *Mean Age in Years & Distribution of Gender by Administration*

Administration	Age in Years					Gender (%)	
	<i>N</i>	<i>M</i>	<i>SD</i>	Min	Max	Male	Female
<b>2004 - 2005</b>	2673	4.61	0.35	3.53	6.20	51.66%	48.34%
<b>2005 - 2006</b>	2642	4.60	0.37	3.02	6.41	48.79%	51.21%
<b>2006 - 2007</b>	2720	4.60	0.35	3.56	6.21	49.82%	50.18%
<b>2007 - 2008</b>	3086	4.59	0.34	3.67	5.99	48.22%	51.78%
<b>2008 - 2009</b>	3041	4.60	0.33	3.69	6.69	49.36%	50.64%
<b>2009 - 2010</b>	2865	4.60	0.33	3.75	6.77	51.83%	48.17%
<b>2010 - 2011</b>	3119	4.61	0.33	3.36	5.98	50.37%	49.63%
<b>2011 - 2012</b>	3054	4.62	0.33	3.06	6.90	49.12%	50.88%
<b>2012 - 2013</b>	2716	4.63	0.32	3.88	6.03	51.10%	48.90%

#### *Grade Testing*

Although the mean age and the proportion of children by gender taking the test has remained relatively consistent across administrations, as can be seen from Table 2, *when* children are taking the assessment has changed substantially over the course of the nine administrations. The number of children testing in Nursery school has declined by 28.52%, and the number of children testing in Pre-Kindergarten has increased by 26.73%, from the 2004-5 to the 2012-13 administrations.

Not only has the proportion of children testing at specific grades levels changed substantially but the total sample size within these grades has also changed. Approximately 50% fewer Nursery school children tested in the 2012-13 administration ( $n = 912$ ), as compared to the 2004-5 ( $n = 1,660$ ) administration. Inversely, 26.93% more Pre-Kindergarten children tested in the 2012-13 administration ( $n = 1,666$ ), as compared to the baseline year ( $n = 925$ ).

Table 2. *Percent of WPPSI-III Test Takers by Grade at Time of Testing*

Administration	Nursery		Pre-Kindergarten		Kindergarten	
	N	Percent	N	Percent	N	Percent
2004 - 2005	1660	62.10%	925	34.61%	88	3.29%
2005 - 2006	1532	57.99%	985	37.28%	125	4.73%
2006 - 2007	1212	44.56%	1385	50.92%	123	4.52%
2007 - 2008	1325	42.94%	1615	52.33%	146	4.73%
2008 - 2009	1255	41.27%	1635	53.77%	151	4.97%
2009 - 2010	1091	38.08%	1626	56.75%	148	5.17%
2010 - 2011	1180	37.83%	1787	57.29%	152	4.87%
2011 - 2012	1095	35.85%	1816	59.46%	143	4.68%
2012 - 2013	912	33.58%	1666	61.34%	138	5.08%

Table 3 provides the mean age in years of the children taking the WPPSI-III across administrations disaggregated by grade at testing. As can be seen, children taking the test in Pre-Kindergarten are, on average, about one to two months older than those children in Nursery school, and children testing in Kindergarten are about five to seven months older than those children testing in Pre-Kindergarten.

Table 3. *Mean Age in Years of WPPSI-III Test Takers across Administrations by Grade at Time of Testing*

Administration	Nursery		Pre-Kindergarten		Kindergarten	
	Mean	SD	Mean	SD	Mean	SD
2004 - 2005	4.58	0.33	4.63	0.35	5.15	0.36
2005 - 2006	4.55	0.34	4.60	0.35	5.21	0.38
2006 - 2007	4.54	0.31	4.60	0.33	5.15	0.33
2007 - 2008	4.53	0.31	4.59	0.32	5.11	0.31
2008 - 2009	4.53	0.30	4.61	0.32	5.10	0.27
2009 - 2010	4.51	0.29	4.62	0.32	5.08	0.31
2010 - 2011	4.51	0.30	4.63	0.32	5.11	0.23
2011 - 2012	4.52	0.29	4.64	0.33	5.10	0.34
2012 - 2013	4.53	0.28	4.63	0.31	5.12	0.27

---

## Across Years, is there Evidence of Trends in WPPSI-III Mean Scaled Scores by Gender or a Child's Grade at Testing?

---

Mean ERB adapted WPPSI-III composite scores (i.e., full scale, verbal, & performance scores) are disaggregated by gender and grade at testing to evaluate if there are any discernable trends in mean scores between groups.

### *Gender*

As can be seen in Figures 1 through 3, female Kindergarten applicants appear to have higher mean full scale, verbal, and performance scores than male Kindergarten applicants across all administrations. Generally, females score about two to three points higher than males on the full scale and approximately one to three points higher than males on the verbal and performance composite scores. Overall, children's mean verbal scores are slightly higher than their performance scores, regardless of gender. Tables A.1 through A.3 in Appendix A provide complete descriptive information disaggregated by gender across the nine administrations.

### *Grade at Testing*

Figures 4 through 6 illustrate trends in mean composite scores disaggregated by children's grade at testing time. As can be seen, there is a substantial difference in mean composite scores across Kindergarten applicants depending on what grade they were in when they took the ERB adapted WPPSI-III. Trends suggest that children testing in Nursery school have higher mean scores on each of the composite scores in comparison to Pre-Kindergarten and Kindergarten grade level applicants, regardless of administration. What is also apparent is that there is very little gain in mean scores across time in children's verbal composite scores; however, there does appear to be small increases in children's mean full scale and performance scores across time at each grade level.

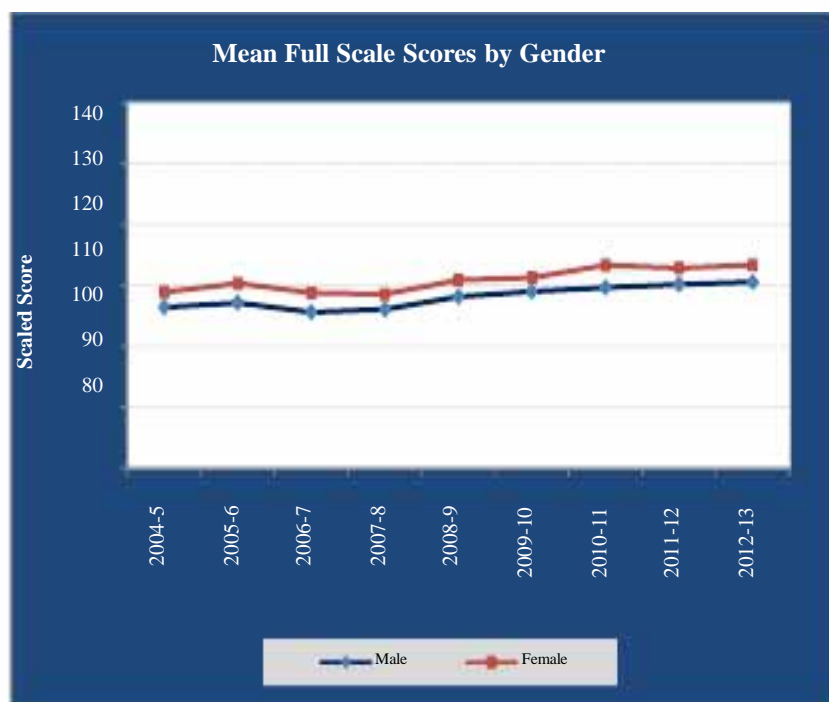


Figure 1. Mean Full Scale Scores by Gender

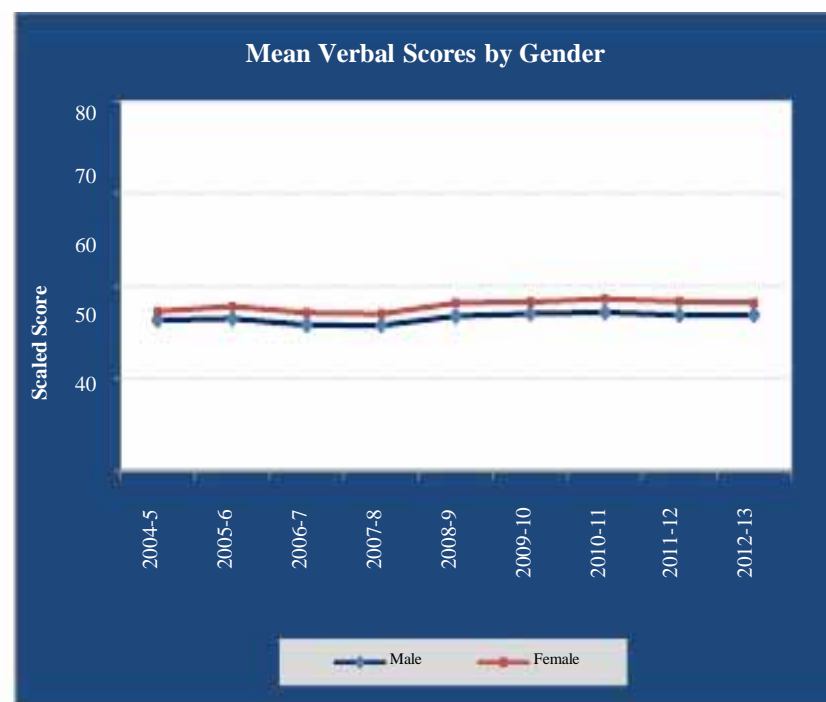


Figure 2. Mean Verbal Scores by Gender

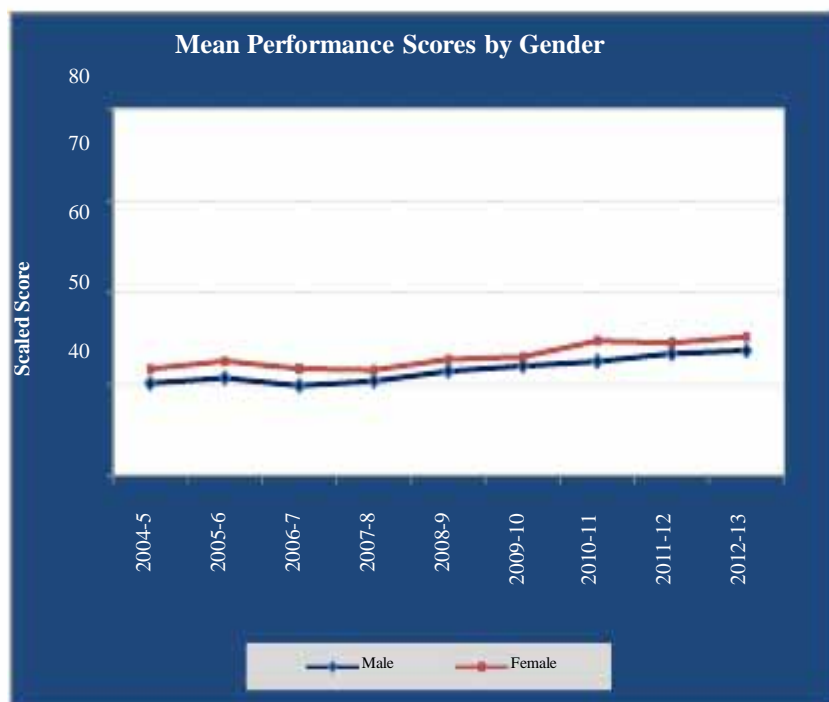


Figure 3. Mean Performance Scores by Gender

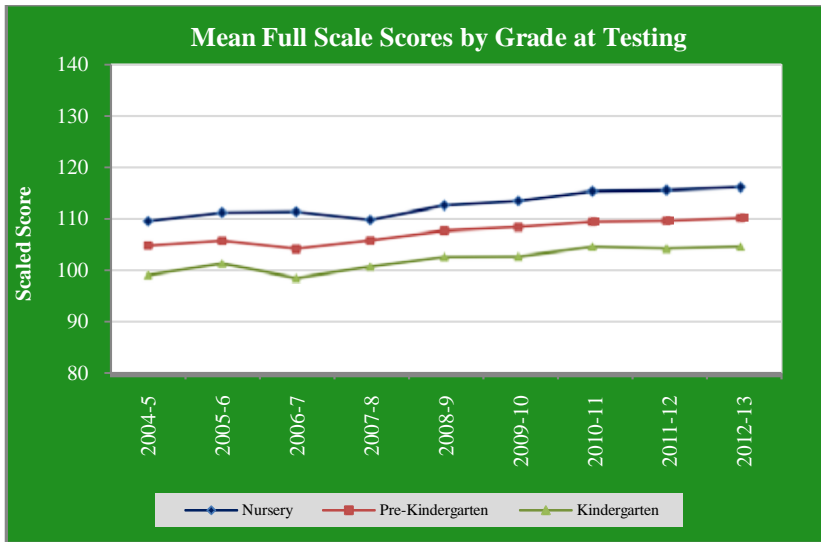


Figure 4. Mean Full Scale Scores by Grade at Testing

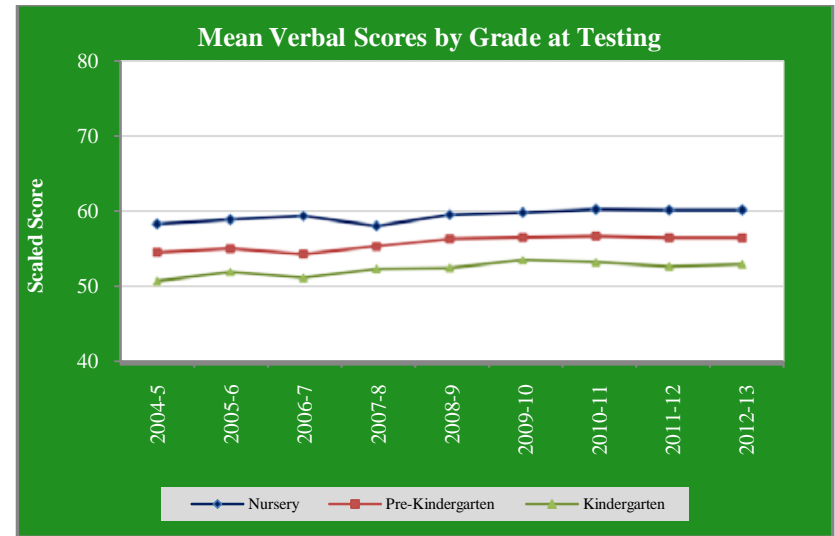


Figure 5. Mean Verbal Scores by Grade at Testing

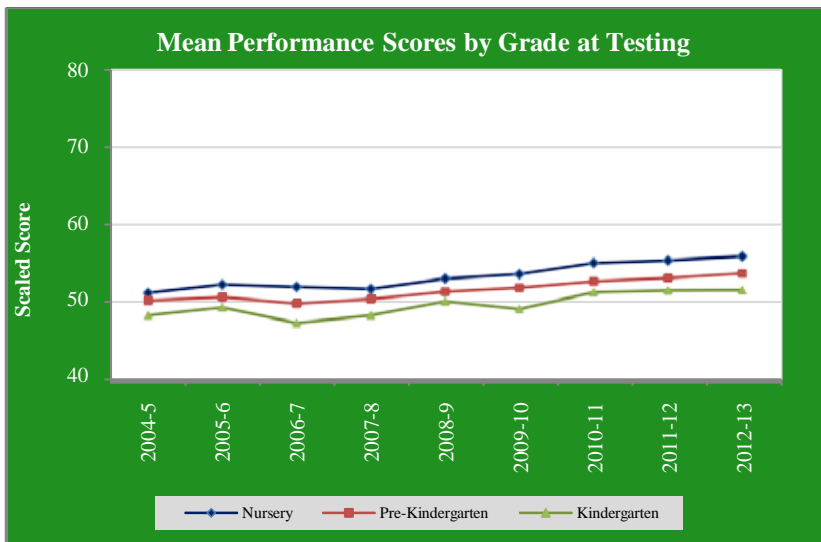


Figure 6. Mean Performance Scores by Grade at Testing

---

## Across Years, is there Evidence of Trends in WPPSI-III Mean Scaled Scores across the Aggregate Sample?

---

Mean ERB adapted WPPSI-III composite scores (i.e., full scale, verbal and performance scores) are evaluated across administrations to assess if there are any discernable trends or patterns in these scores across time. Results are described for each composite score using mean scaled scores, scaled score standard deviations, and standardized mean differences calculate between each consecutive year.

### *Mean Scores Across Administrations*

Descriptive statistics for the ERB adapted WPPSI-III composite scaled scores by administration are summarized in Table 4. As can be seen, the full scaled scores range between  $M = 107.61$  ( $SD = 13.54$ ) to  $M = 111.91$  ( $SD = 14.23$ ), verbal scores range between  $M = 56.38$  ( $SD = 8.45$ ) and  $M = 57.86$  ( $SD = 8.07$ ) and performance scaled scores range between  $M = 50.85$  ( $SD = 7.30$ ) and  $M = 54.40$  ( $SD = 8.03$ ) across the nine administrations. Mean score trends across time are also illustrated in figures 7 and 8 for the full scale and the verbal/performance scaled scores, respectively.

Evaluation of scores indicates that the full scale scores have increased linearly by approximately four scaled score points from the 2004–5 to the 2012-13 administrations. The increase in mean full scale scores, however, has increased very little since the 2009-10 administration (1.83 scaled score points). Unlike the strict linear trend that is observed when evaluating the mean full scale scores across time, the verbal and performance scores appear to be much more variable across administrations. Like the full scaled scores, the performance composite mean scores have also increased approximately four scaled score points from the baseline 2004-2005 data. These mean scores however, like the full scale means, have increased very gradually over time. Evaluation of the verbal mean scores across time, however, indicates that these scores vary in magnitude and direction, depending on the administration year. There is evidence that the 2012-13 verbal mean scores have declined by less than a scale score point since the 2004-5 administration.

Table 4. *Mean WPPSI-III Composite Scores Across Administrations*

Administration	Full Scale			Verbal			Performance		
	<i>N</i>	Mean	SD	<i>N</i>	Mean	SD	<i>N</i>	Mean	SD
2004 - 2005	2673	107.61	13.54	2673	56.76	8.54	2673	50.85	7.30
2005 - 2006	2642	108.76	13.92	2642	57.13	8.69	2642	51.62	7.45
2006 - 2007	2720	107.19	14.62	2720	56.43	9.13	2720	50.76	7.84
2007 - 2008	3086	107.34	13.98	3086	56.38	8.45	3086	50.96	7.76
2008 - 2009	3041	109.52	13.20	3041	57.45	8.06	3041	52.07	7.35
2009 - 2010	2865	110.08	13.21	2865	57.61	7.95	2865	52.47	7.49
2010 - 2011	3119	111.45	13.84	3119	57.86	8.07	3119	53.59	7.81
2011 - 2012	3054	111.52	13.96	3054	57.60	8.08	3054	53.93	7.96
2012 - 2013	2716	111.91	14.23	2716	57.51	8.32	2716	54.40	8.03

Three things of major importance are highlighted in the mean score trends presented in Table 4 and Figures 7 and 8. First, both the full scale and the performance scores remain quite stable from the 2010-11 administration. There is very little increase in mean scores during this three-year time period for either of these composite scores (i.e., less than .5 scale score points). Second, verbal mean scores over administrations are remarkably stable and deviate by about only one scale score point from year to year. As the full scale composite score is the summation of both the verbal and performance scores, this finding highlights that the fluctuation seen in the full scale scores is directly attributable to the fluctuation in the performance scores.

Lastly, and most importantly, evaluation of mean scores at the global level indicates that the children taking the ERB adapted WPPSI-III in the admission process are much more able than the standardized sample used for the WPPSI-III clinical assessment. Across all administrations, mean full scale scores are approximately 10 points higher than those of the national normative sample ( $M = 100$ ,  $SD = 15$ ). This strongly indicates that the ERB sample is much different from the national standardization sample. Therefore, using nationally reference percentile ranks may result in artificially inflated relative standings for the ERB sample, regardless of year.



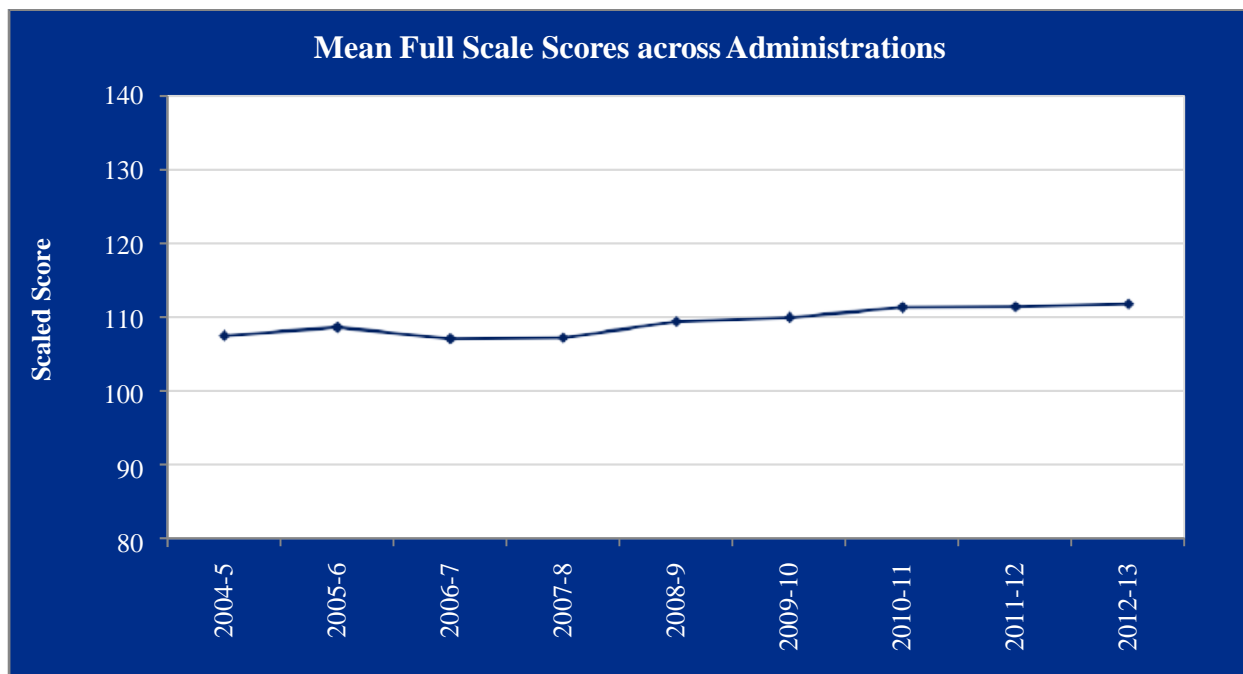


Figure 7. Mean Full Scale Scores Across Administrations

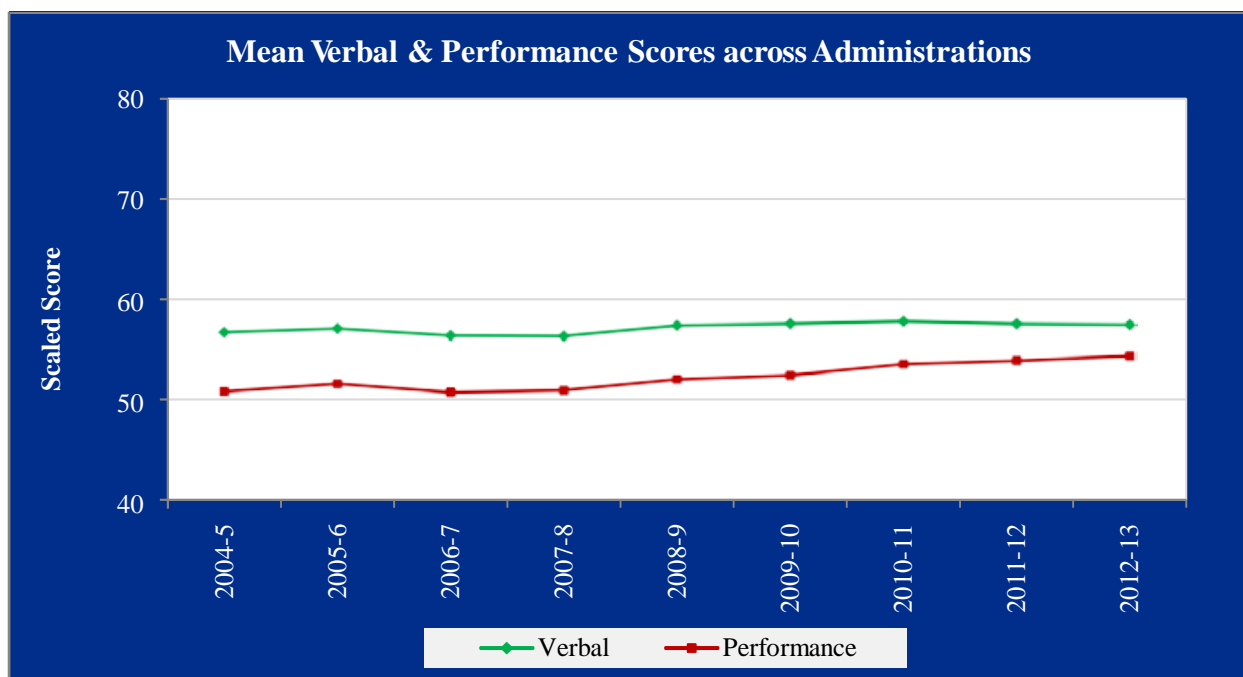


Figure 8. Mean Verbal and Performance Scores Across Administrations

## Have the WPPSI-III Mean Scores Increased Substantially Over Administrations?

One major area of concern for the users of the ERB adapted WPPSI-III is that the scaled scores may be increasing substantially over recent years due to the availability of test preparatory materials. As Table 4 above has indicated, although there do appear to be gains in scores across years in the full scale and performance composite scores, these increases have been relatively gradual over time. To assess if these increases are substantial enough to be of practical significance, score gains across consecutive administrations are evaluated.

Standardized mean differences, or effect sizes, for each of the mean ERB adapted WPPSI-III composite scores are calculated across pairs of consecutive years. Table 5 provides the mean gain scores and effect sizes for the full scale, verbal, and performance composite scores, respectively. As can be seen, both the gain scores and the magnitudes of the effect sizes for each composite score fluctuate from year to year. All effect sizes are small, suggesting that year-to-year mean fluctuations are not of practical significance.

Table 5. *Mean Gain Scores and Standardized Mean Differences of WPPSI-III Composite Scores Across Administrations*

Administrations		Full Scale		Verbal		Performance	
Time 1	Time 2	Gain Score	ES	Gain Score	ES	Gain Score	ES
2004-5	2005-6	1.15	<b>0.08</b>	0.37	<b>0.04</b>	0.77	<b>0.10</b>
2005-6	2006-7	-1.56	<b>-0.11</b>	-0.70	<b>-0.08</b>	-0.86	<b>-0.11</b>
2006-7	2007-8	0.14	<b>0.01</b>	-0.06	<b>-0.01</b>	0.20	<b>0.03</b>
2007-8	2008-9	2.18	<b>0.16</b>	1.08	<b>0.13</b>	1.11	<b>0.15</b>
2008-9	2009-10	0.56	<b>0.04</b>	0.16	<b>0.02</b>	0.39	<b>0.05</b>
2009-10	2010-11	1.37	<b>0.10</b>	0.25	<b>0.03</b>	1.12	<b>0.15</b>
2010-11	2011-12	0.07	<b>0.01</b>	-0.27	<b>-0.03</b>	0.34	<b>0.04</b>
2011-12	2012-13	0.39	<b>0.03</b>	-0.08	<b>-0.01</b>	0.47	<b>0.06</b>

\*Note: ES = Effect size (Standardized Mean Difference). ES = .2 is small, ES = .3 is medium, and ES = .5 is large.

---

## Across Years, Is There Evidence of an Upward Trend in the Proportion of Children Scoring at the 90<sup>th</sup> and the 98<sup>th</sup> Percentiles on the WPPSI-III?

---

### *Full Scale Score Distributions*

Figures B.1 through B.3 in Appendix B represent frequency distributions for the full scale scores across each administration. Full scale scores of 106 and 121 yield the 90<sup>th</sup> and 98<sup>th</sup> percentile ranks, respectively, within each year. These cut scores are indicated in the figures via a dashed black line (90<sup>th</sup> percentile) and a dashed red line (98<sup>th</sup> percentile) to visually depict the number of applicants scoring both below and above these cut scores.

Inspection of the full scale distributions indicate that the sample taking the ERB adapted WPPSI-III is relatively similar across all nine administrations. Not surprisingly, given this samples higher mean score trend across years, the distributions reflect that more children are scoring at the upper end of the distribution, regardless of which administration is evaluated.

Table 6 highlights the percentage of children who score at or above both the 90<sup>th</sup> and 98<sup>th</sup> percentile ranks by administration. Across all administrations, a vast majority of children score at or above the 90<sup>th</sup> and 98<sup>th</sup> percentile rank in comparison to the nationally standardized normative sample. That being said, there are only small to moderate increases in the proportion of children scoring at or above these percentile ranks across consecutive years.

More specifically, the proportion of children scoring at or above the 90<sup>th</sup> percentile rank increases only slightly from the 2009-10 to 2012-13 years (1.95%) and marginally from the 2004-5 to 2012-13 years (8.67%). More modest increases, however, are seen at the 98<sup>th</sup> percentile (2009-10 to 2012-13, 7%; 2004-5 to 2012-13, 11.54%). Figures 9 and 10 illustrate trends associated with the percentage of children scoring at or above the 90<sup>th</sup> and 98<sup>th</sup> percentiles across years.

Table 6. *Percentage of Children Scoring At or Above the 90<sup>th</sup> and 98<sup>th</sup> Percentiles – Full Scale*

Administration	Percentile	
	90th	98th
2004-5	61.73%	17.62%
2005-6	64.35%	20.74%
2006-7	60.40%	18.13%
2007-8	60.34%	17.76%
2008-9	67.25%	20.72%
2009-10	68.45%	22.16%
2010-11	70.25%	27.06%
2011-12	70.27%	27.37%
2012-13	70.40%	29.16%

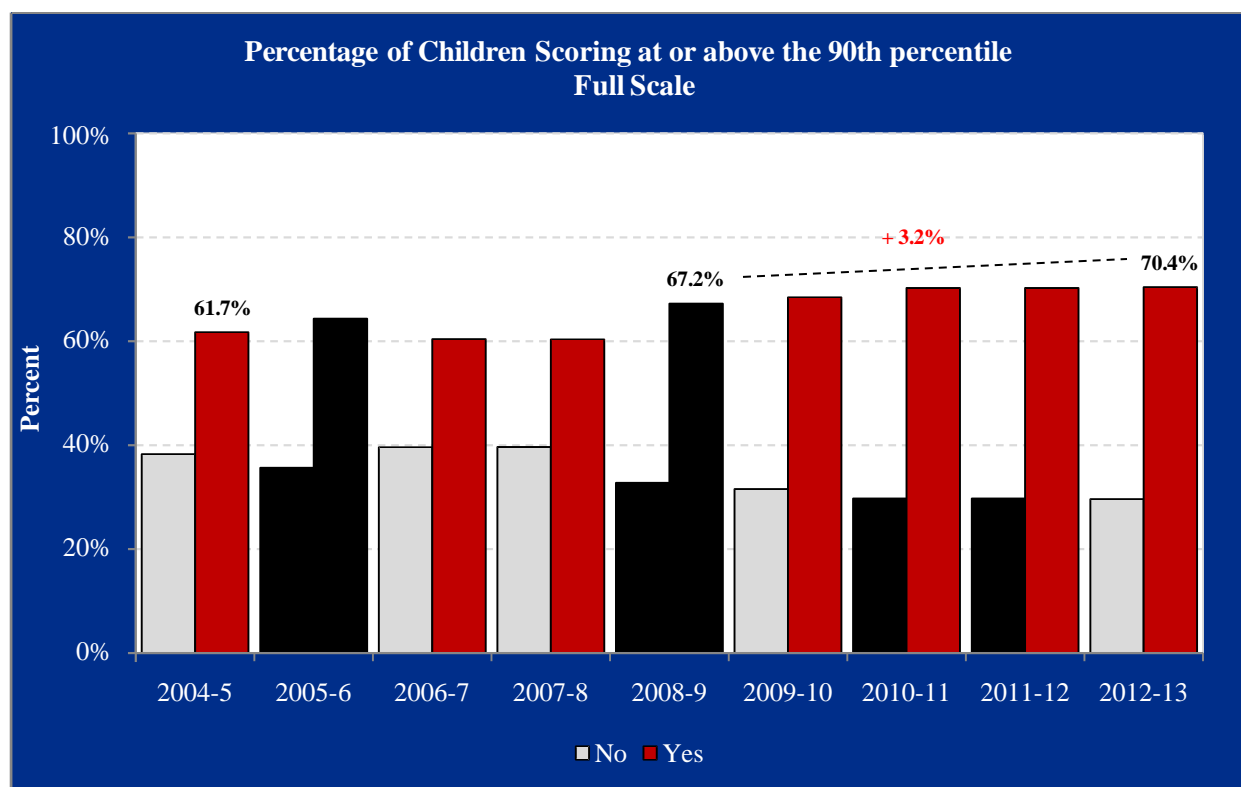


Figure 9. *Percentage of Children At or Above the 90<sup>th</sup> Percentile – Full Scale*

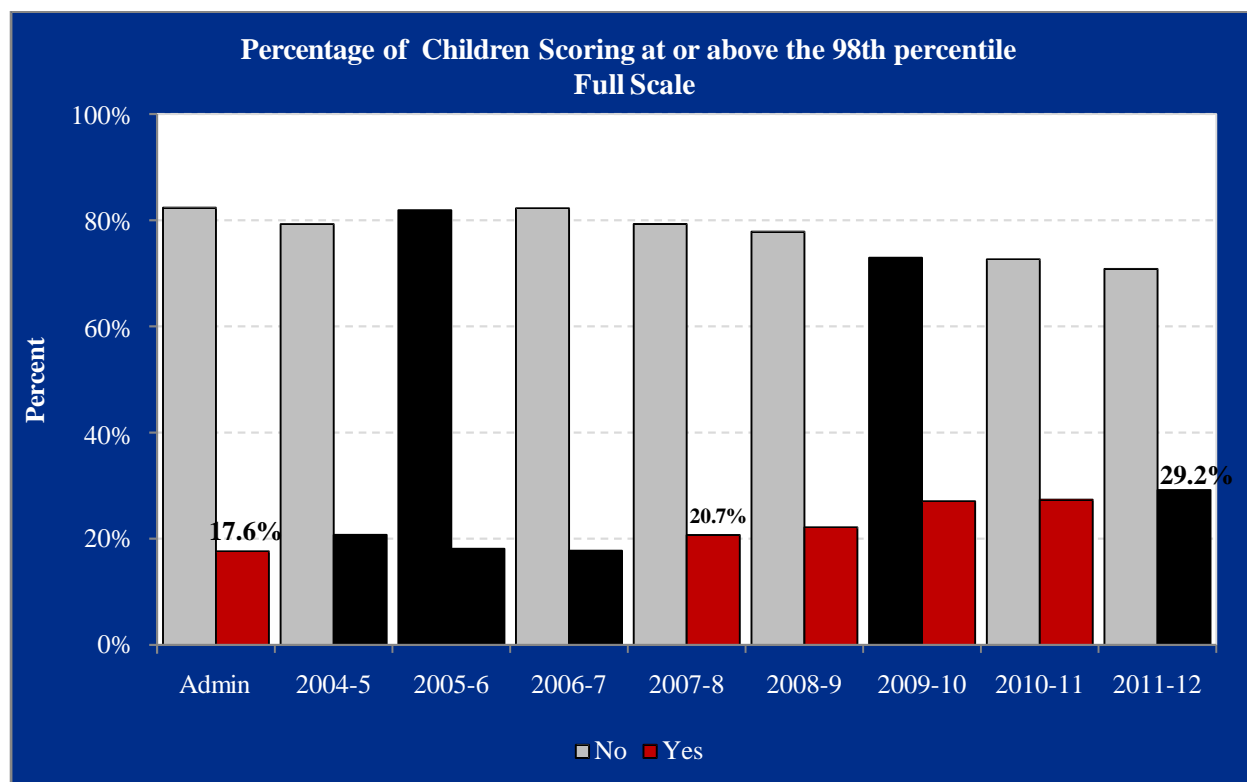


Figure 10. Percentage of Children At or Above the 90<sup>th</sup> Percentile – Full Scale

#### *Verbal and Performance Scale Score Distributions*

Figures C.1 through C.3 in Appendix C, and figures D.1 through D.3 in Appendix D, illustrate the frequency distributions across administrations for the verbal and performance scores, respectively.

For each administration, verbal scaled scores of 54 and 61, and performance scaled scores of 52 and 59, yield the 90<sup>th</sup> and 98<sup>th</sup> percentile rank, respectively. Again, as was indicated for the full scale score distributions, the 90<sup>th</sup> and 98<sup>th</sup> percentile cut scores are indicated via dashed black and red lines in the figures.

As was observed when evaluating the full scale score distributions, both the verbal and the performance scaled score distributions appear relatively similar across years, with more children scoring at the upper end of the distributions. It can be seen that, in general, children's performance scores are somewhat lower than mean verbal scores, especially in earlier administrations. This was also found when evaluating the mean statistics above.

Table 7 includes the percentage of children scoring at or above both the 90<sup>th</sup> and 98<sup>th</sup> percentiles for both the verbal and the performance scores by administration. Although the percentage of children at or above the 90<sup>th</sup> percentile on the verbal composite increases by about 2% across the nine administrations, what can also be seen is that the proportion of children at the 90<sup>th</sup> percentile actually decreases by approximately 2.5% between the 2007-8 and 2012-13 administrations. At the 98th percentile the trend is similar; however, there appears to be more variation in the number of children scoring at or above this percentile rank across time, with the proportion of children decreasing slightly across the more recent administrations.

Evaluation of the percentage of children scoring at these percentiles on the performance score tells a much different story, however. The percentage of children at or above the 90<sup>th</sup> and 98<sup>th</sup> percentile in the 2012-13 administration increases by 9.4% and 13.4% when compared to the 2007-8 administration, and 16% and 18% when compared to the 2004-5 administration. Figures 11 and 12 illustrate these trends across administrations for the verbal and performance scores, respectively.

Table 7. *Percentage of Children Scoring At or Above the 90<sup>th</sup> and 98<sup>th</sup> Percentiles on the Verbal and Performance Composite Scores*

Administration	Percentiles			
	Verbal		Performance	
	90th	98th	90th	98th
2004-5	67.27%	37.37%	48.15%	14.78%
2005-6	68.36%	39.93%	52.88%	18.13%
2006-7	65.26%	36.99%	47.43%	16.88%
2007-8	65.98%	35.13%	48.67%	17.14%
2008-9	71.62%	39.56%	54.98%	19.40%
2009-10	72.18%	39.93%	56.51%	23.07%
2010-11	74.09%	41.49%	62.78%	27.89%
2011-12	71.15%	40.01%	62.54%	30.22%
2012-13	69.22%	40.10%	64.36%	32.77%

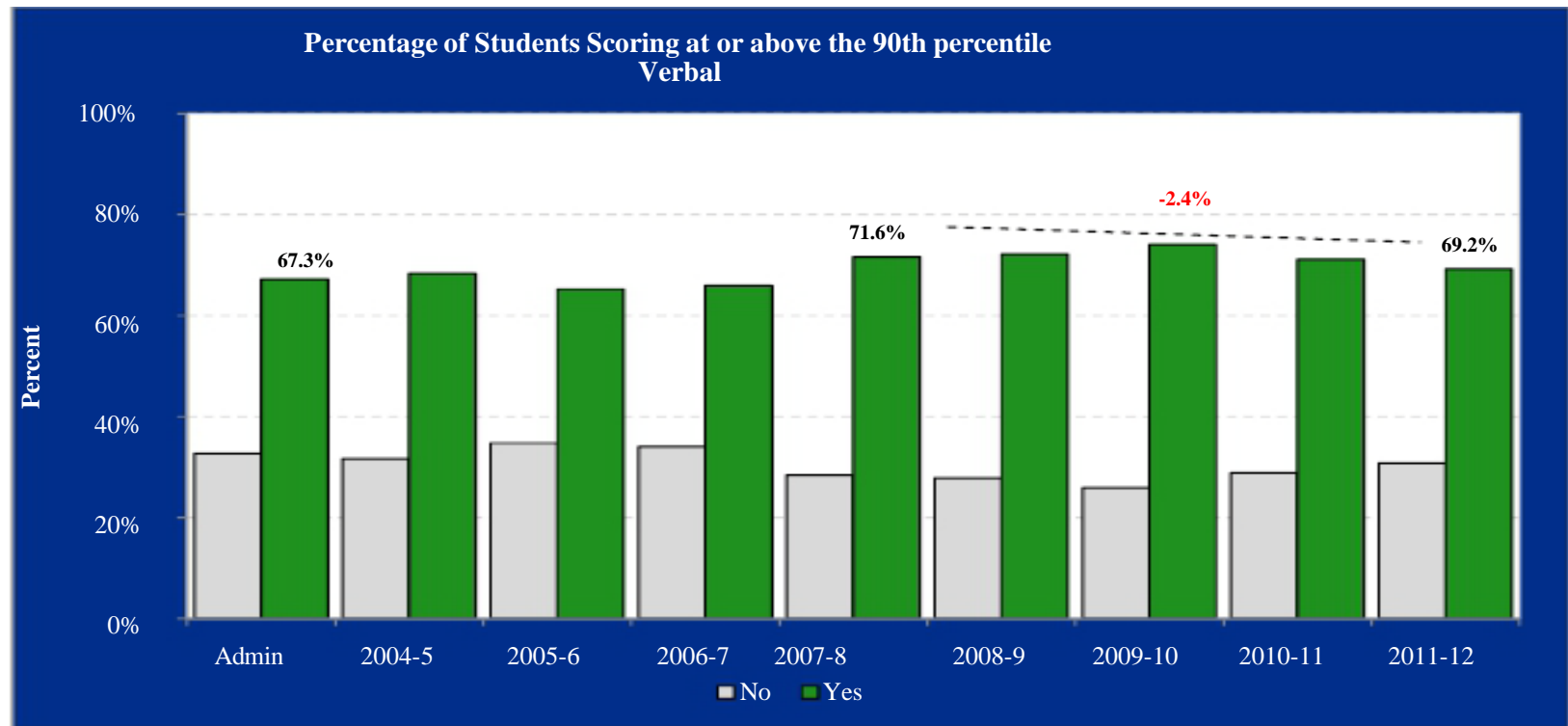


Figure 11. Percentage of Children at or above the 90<sup>th</sup> Percentile – Verbal

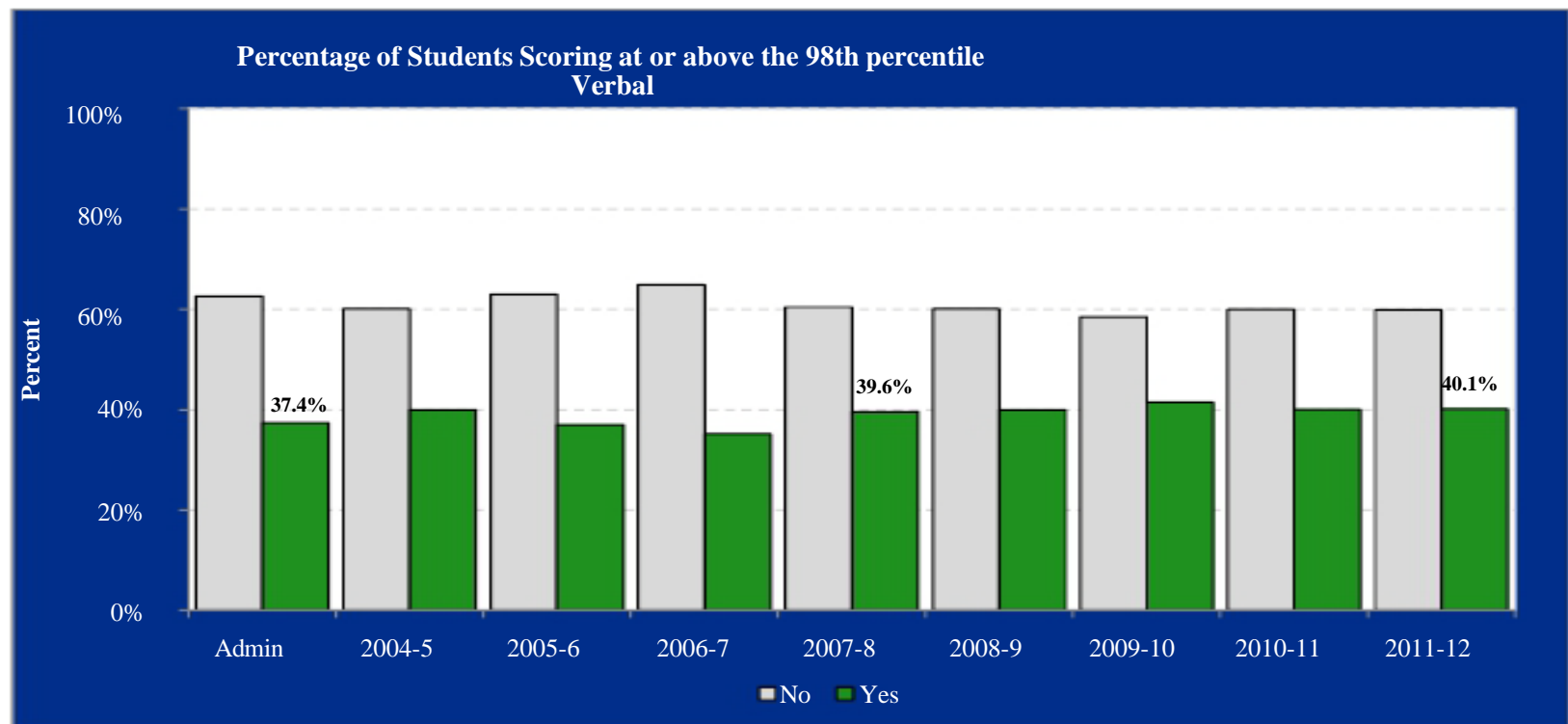


Figure 12. Percentage of Children at or above the 98<sup>th</sup> Percentile – Verbal

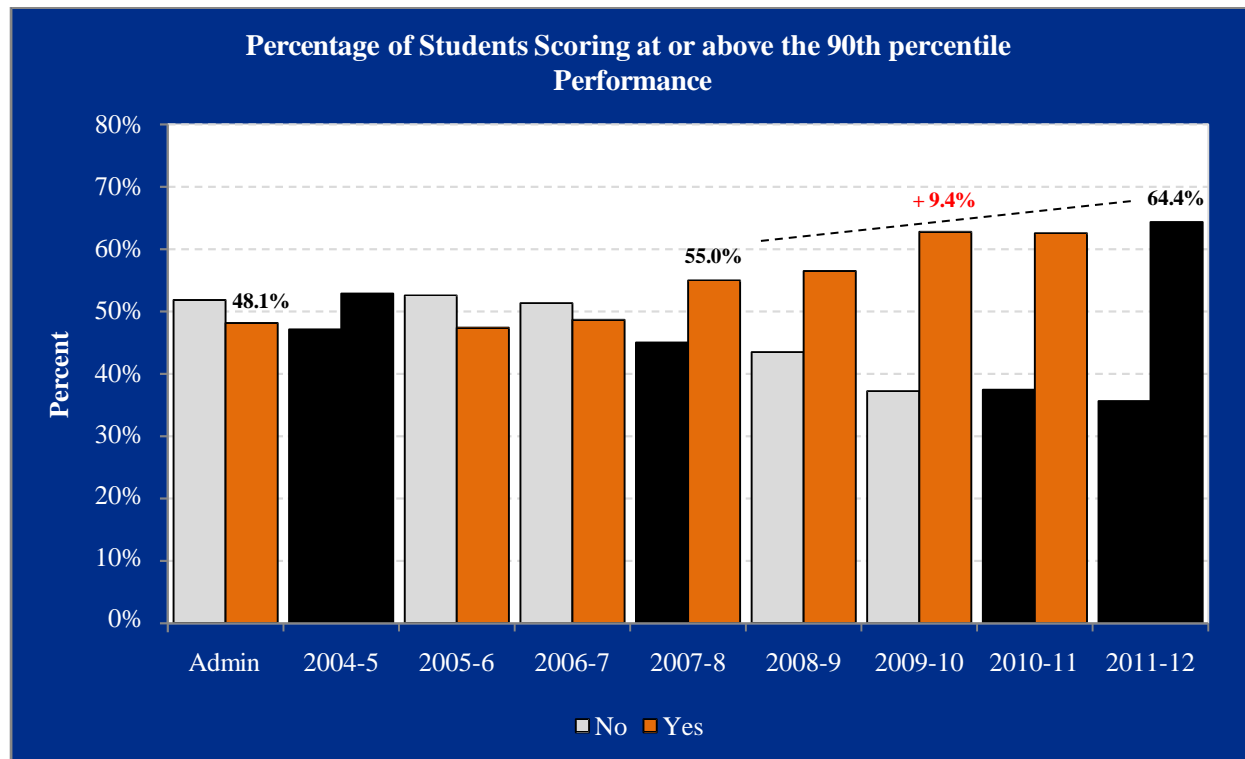


Figure 13. Percentage of Children At or Above the 90<sup>th</sup> Percentile – Performance

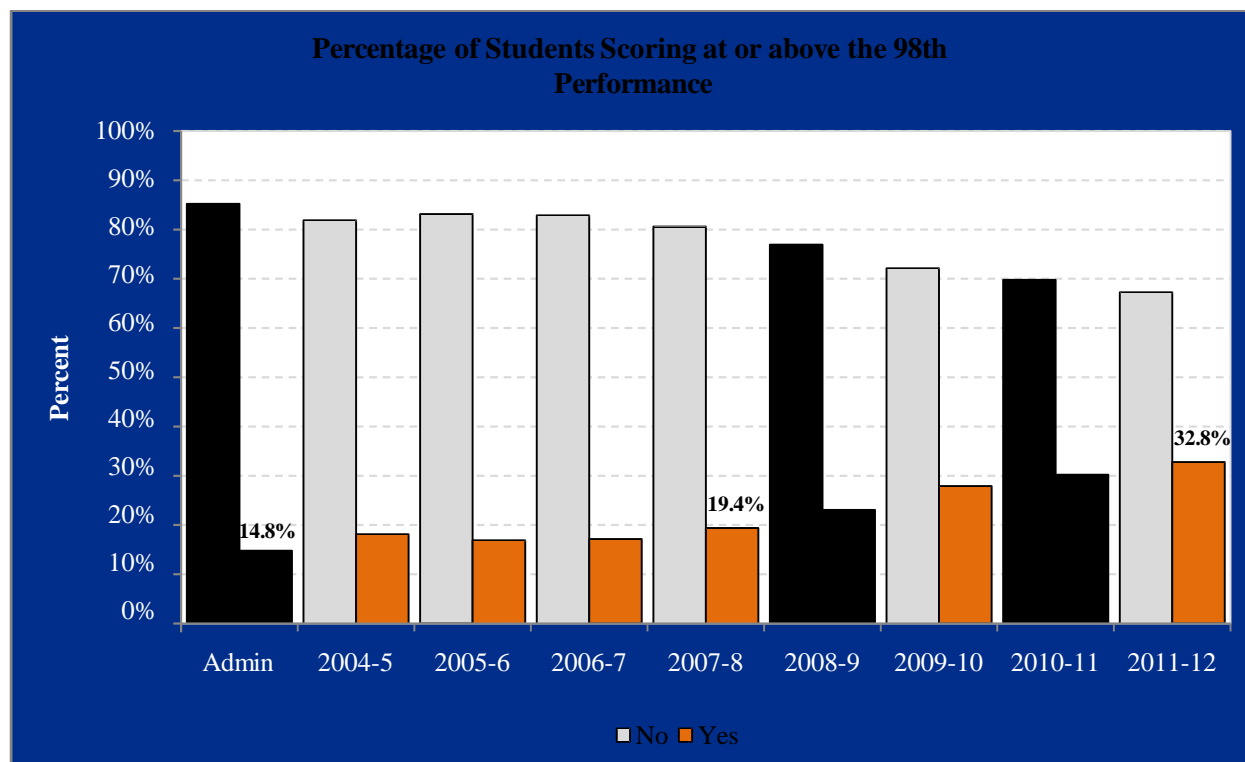


Figure 14. Percentage of Children At or Above the 98<sup>th</sup> Percentile – Performance



---

## Does the Proportion of Children Scoring at the 90<sup>th</sup> and 98<sup>th</sup> Percentiles on the WPPSI-III Change Significantly Over Time?

---

To assess if increases in the proportion of children scoring at or above the 90<sup>th</sup> and 98<sup>th</sup> percentile are significant across years, the  $\chi^2$  test is used. In essence, the  $\chi^2$  test asks: ‘Does the number of children that score at or above the 90<sup>th</sup> or 98<sup>th</sup> percentile differ significantly from what we would expect?’

As can be seen from Table 8, throughout the 2008–9 and the 2012-13 administrations, the proportion of children scoring at or above the 90<sup>th</sup> percentile on the full scale is not found to be significantly different than what would be expected across this administration time frame. However, there is a significant difference in the proportion of children scoring at these percentiles at different time points between the earlier administrations (i.e., 2004-5 to 2008-9). These differences vary in direction, with scores either increasing or decreasing depending on the administration period. These same fluctuations and directional changes are also seen when looking at the proportion of children scoring at or above the 98<sup>th</sup> percentile.

Tables 9 and 10 highlight the proportion of children scoring at the 90<sup>th</sup> and 98<sup>th</sup> percentiles for the verbal and performance scores, respectively, and the resulting  $\chi^2$  statistics across consecutive years. As can be seen, the proportion of children who score at or above the 90<sup>th</sup> percentile on the verbal composite significantly decreases across the 2010-11 and 2011-12 years. In addition, there are no significant differences between the proportions of children who score at or above the 98<sup>th</sup> percentiles across consecutive years are beginning with the 2008-9 administration.

The performance score shows a much different pattern (Table 11). More children are scoring at the 90<sup>th</sup> percentile in 2010-11 when compared to the 2009-10 year, and the proportion of children scoring at or above the 98<sup>th</sup> percentile are significantly more than what we would expect across consecutive administrations beginning in the 2007-8 year.

Table 8.  $\chi^2$  Difference test of Proportions across Administrations – Full Scale

Full Scale													
90th Percentile							98th Percentile						
Time 1	%	Time 2	%	$\chi^2$	Sig.	Direction of Change	Time 1	%	Time 2	%	$\chi^2$	Sig.	Direction of Change
2004-5	61.7	2005-6	64.3	3.90	*	+	2004-5	17.6	2005-6	20.7	8.35	**	+
2005-6	64.3	2006-7	60.4	8.86	**	-	2005-6	20.7	2006-7	18.1	5.87	*	-
2006-7	60.4	2007-8	60.3	0.00	ns	-	2006-7	18.1	2007-8	17.8	0.13	ns	-
2007-8	60.3	2008-9	67.2	31.65	**	+	2007-8	17.8	2008-9	20.7	8.64	**	+
2008-9	67.2	2009-10	68.4	0.97	ns	+	2008-9	20.7	2009-10	22.2	1.84	ns	+
2009-10	68.4	2010-11	70.2	2.28	ns	+	2009-10	22.2	2010-11	27.1	19.23	**	+
2010-11	70.2	2011-12	70.3	0.00	ns	+	2010-11	27.1	2011-12	27.4	0.08	ns	+
2011-12	70.3	2012-13	70.4	0.01	ns	+	2011-12	27.4	2012-13	29.2	2.26	ns	+

Note: \* $p < .05$ ; \*\* $p < .001$ ; ns = Not Significant; + = Positive Gain; - = Negative Gain

Table 9.  $\chi^2$  Difference test of Proportions across Administrations – Verbal

Verbal													
90th Percentile							98th Percentile						
Time 1	%	Time 2	%	$\chi^2$	Sig.	Direction of Change	Time 1	%	Time 2	%	$\chi^2$	Sig.	Direction of Change
2004-5	67.3	2005-6	68.4	0.73	ns	+	2004-5	37.4	2005-6	39.9	3.67	*	+
2005-6	68.4	2006-7	65.3	5.81	*	-	2005-6	39.9	2006-7	37.0	4.92	*	-
2006-7	65.3	2007-8	66.0	0.33	ns	+	2006-7	37.0	2007-8	35.1	2.17	ns	-
2007-8	66.0	2008-9	71.6	22.73	**	+	2007-8	35.1	2008-9	39.6	12.86	**	+
2008-9	71.6	2009-10	72.2	0.23	ns	+	2008-9	39.6	2009-10	39.9	0.09	ns	+
2009-10	72.2	2010-11	74.1	2.78	ns	+	2009-10	39.9	2010-11	41.5	1.50	ns	+
2010-11	74.1	2011-12	71.2	6.72	*	-	2010-11	41.5	2011-12	40.0	1.39	ns	-
2011-12	71.2	2012-13	69.2	2.57	ns	-	2011-12	40.0	2012-13	40.1	0.00	ns	+

Note: \* $p < .05$ ; \*\* $p < .001$ ; ns = Not Significant; + = Positive Gain; - = Negative Gain

Table 10.  $\chi^2$  Difference test of Proportions across Administrations – Performance

Performance													
90th Percentile							98th Percentile						
Time 1	%	Time 2	%	$\chi^2$	Sig.	Direction of Change	Time 1	%	Time 2	%	$\chi^2$	Sig.	Direction of Change
2004-5	48.1	2005-6	52.9	11.88	**	+	2004-5	14.8	2005-6	18.1	10.87	**	+
2005-6	52.9	2006-7	47.4	15.92	**	-	2005-6	18.1	2006-7	16.9	1.46	ns	-
2006-7	47.4	2007-8	48.7	0.90	ns	+	2006-7	16.9	2007-8	17.1	0.07	ns	+
2007-8	48.7	2008-9	55.0	24.43	**	+	2007-8	17.1	2008-9	19.4	5.24	*	+
2008-9	55.0	2009-10	56.5	1.40	ns	+	2008-9	19.4	2009-10	23.1	11.90	**	+
2009-10	56.5	2010-11	62.8	24.39	**	+	2009-10	23.1	2010-11	27.9	18.23	**	+
2010-11	62.8	2011-12	62.5	0.04	ns	-	2010-11	27.9	2011-12	30.2	4.06	*	+
2011-12	62.5	2012-13	64.4	2.05	ns	+	2011-12	30.2	2012-13	32.8	4.32	*	+

Note: \* $p < .05$ ; \*\* $p < .001$ ; ns = Not Significant; + = Positive Gain; - = Negative Gain

---

## **Across Years, Is There Evidence of an Upward Trend in the Proportion of Children Scoring at the 90<sup>th</sup> and the 98<sup>th</sup> Percentiles on the WPPSI-III Based on the ERB Empirical Samples?**

---

As was seen in the descriptive statistics presented above, as well as from the score distributions of children taking the ERB adapted WPPSI-III for admission, the ERB samples, on average, score about 10 points higher than the nationally normed sample with respect to the full scale score. The substantially larger mean scores in the ERB samples implicitly suggest that the children taking the WPPSI-III during the admission process may not be directly comparable to the national normative sample. This becomes even more apparent when evaluating the proportion of children at and above the 90<sup>th</sup> percentile, as a large proportion of children are indeed, scoring at or above the 90<sup>th</sup> and 98<sup>th</sup> percentile — regardless of administration.

To gain a fuller understanding as to whether the proportion of children scoring at the upper end of the distribution is in fact changing, empirical percentile rankings by administration for each WPPSI- III composite score are calculated.

### *Full Scale Percentile Rank Distributions based on the Empirical Samples*

Figures 15 and 16 illustrate the proportion of children scoring at or above the 90<sup>th</sup> and 98<sup>th</sup> percentiles, respectively, on the WPPSI-III full scale score. As can be seen, the empirical percentiles paint a completely different picture from what was found above. Across the nine administrations, approximately 11% of the children are scoring at or above the 90<sup>th</sup> percentile and approximately 3% scoring at or above the 98<sup>th</sup> percentile. This is in stark contrast to the large proportion of children scoring at these percentile ranks using the nationally normed percentile ranks.

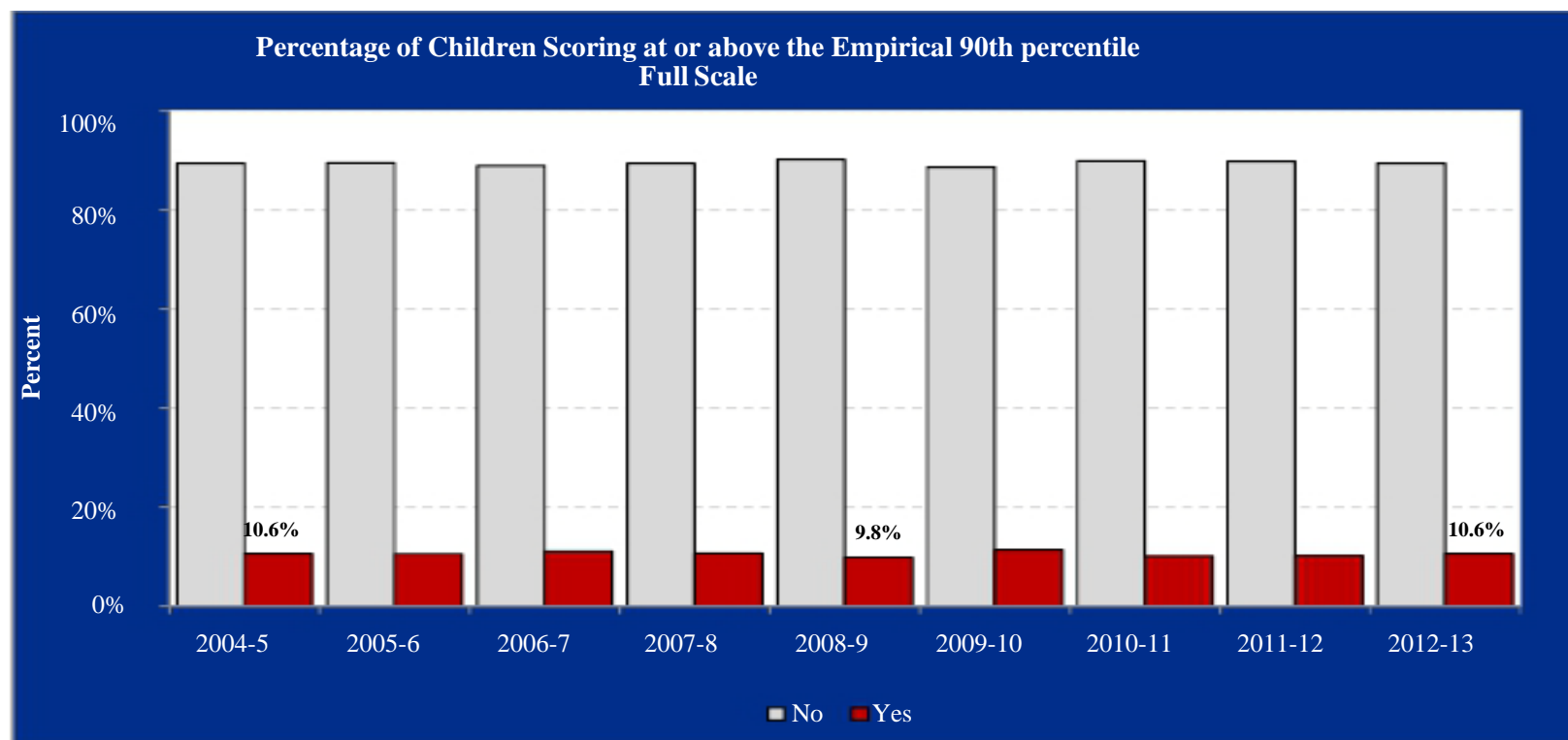


Figure 15. Percentage of Children at or above the 90<sup>th</sup> Empirical Percentile – Full Scale

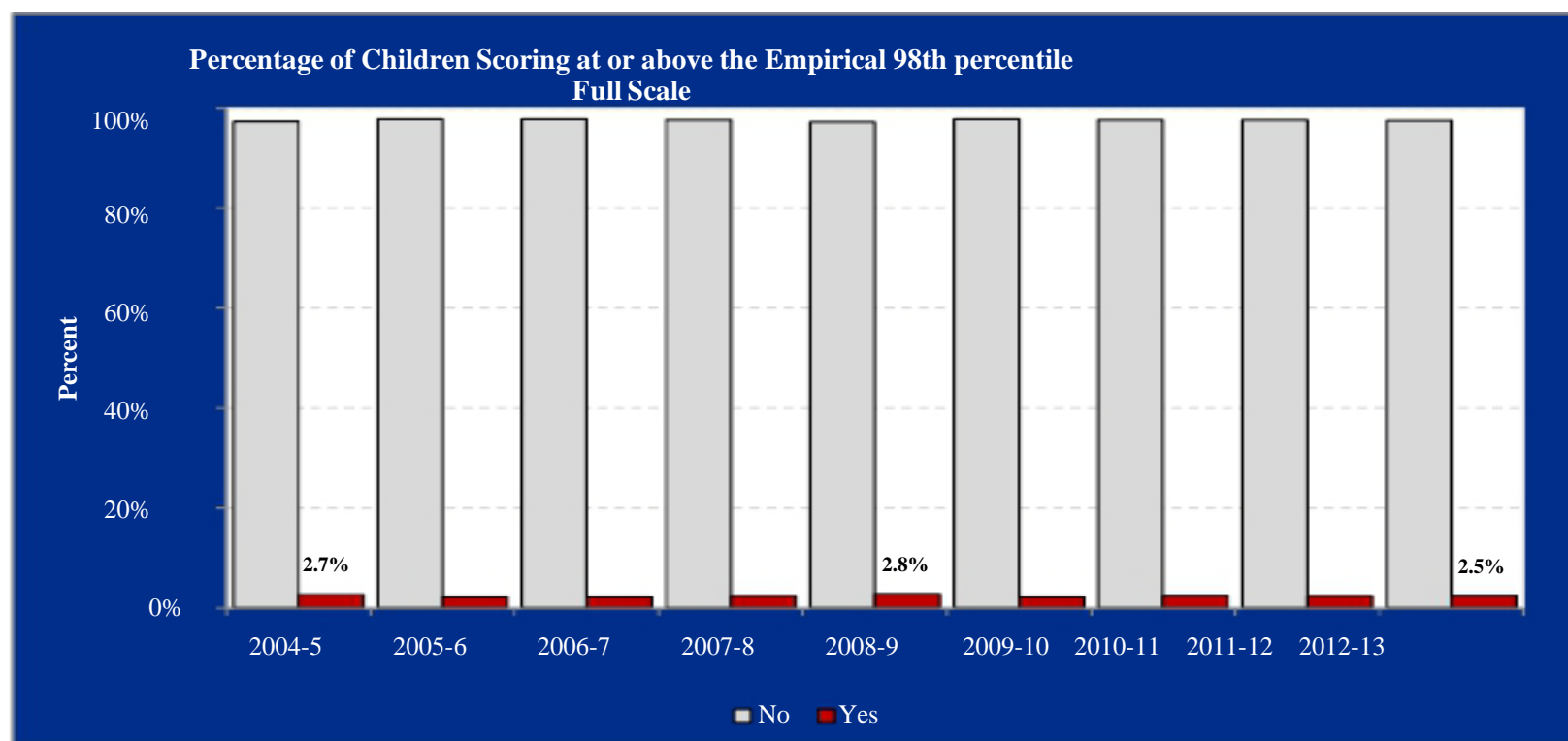
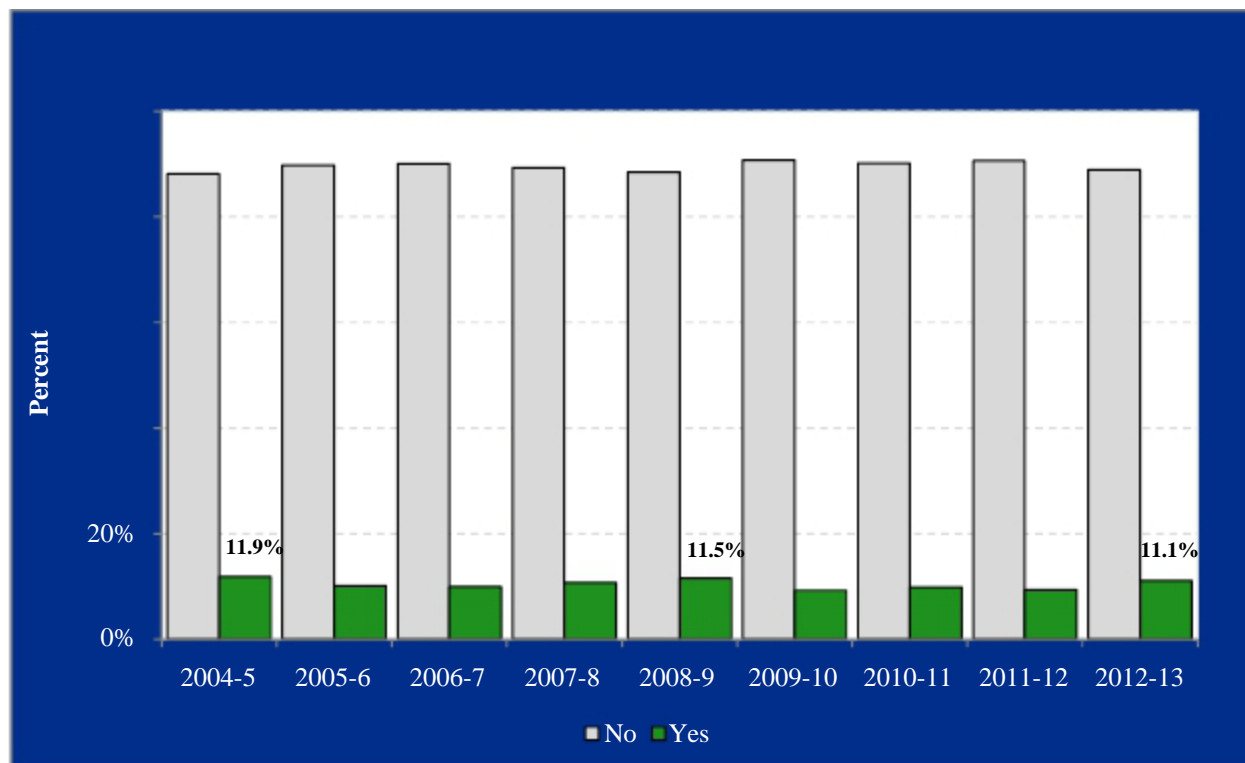


Figure 16. Percentage of Children at or above the 98<sup>th</sup> Empirical Percentile – Full Scale

### *Verbal and Performance Percentile Rank Distributions Based on the Empirical Samples*

Using empirical percentile rankings based on the sample data for the verbal and performance scores illustrates the same trends. As can be seen in Figures 17 through 20, fewer than 12% of children are scoring at the 90<sup>th</sup> percentile on both the verbal and performance composites when the empirical sample is used to calculate percentiles, and fewer than 3% are scoring at or above the 98<sup>th</sup> percentile. Generally, the proportion of children scoring at or above both the 90<sup>th</sup> and 98<sup>th</sup> percentile ranks actually decreases between the 2004-5 and 2012-13 administrations across both the verbal and performance measures.



*Figure 17. Percentage of Children at or above the 90<sup>th</sup> Empirical Percentile – Verbal*

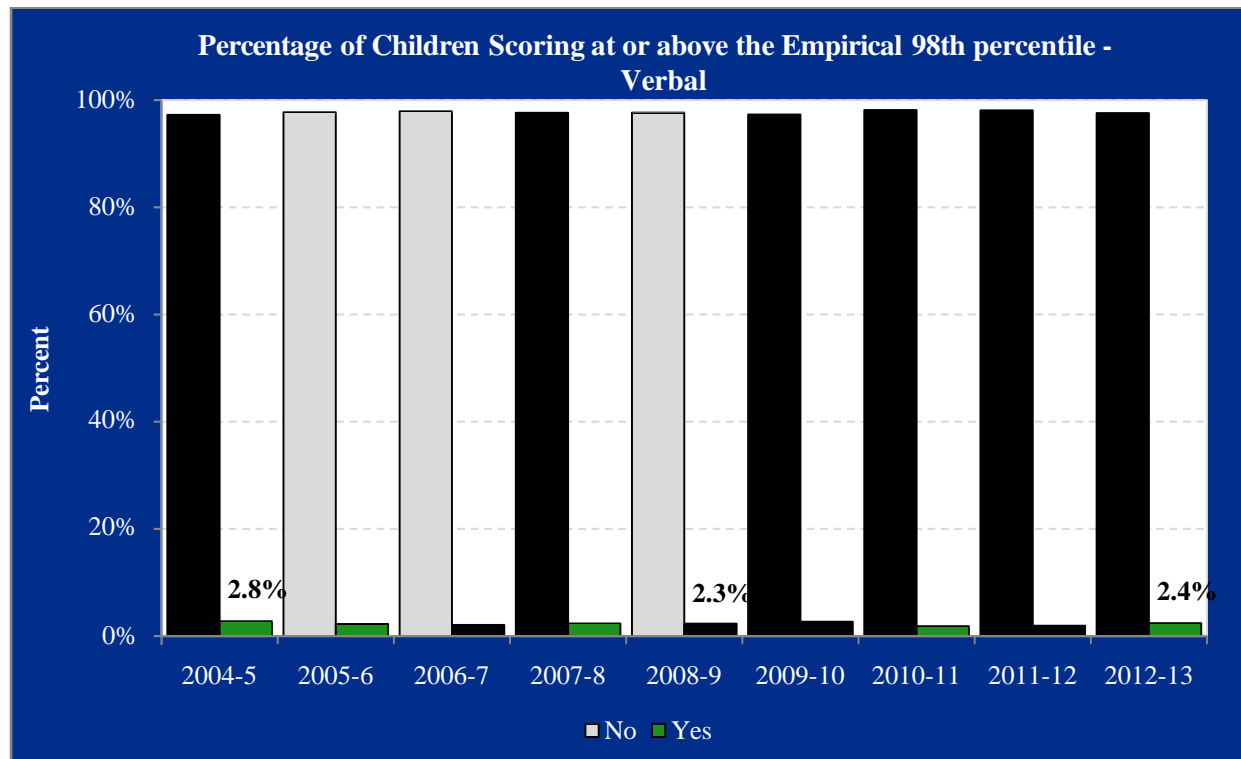


Figure 18. Percentage of Children At or Above the 98<sup>th</sup> Empirical Percentile – Verbal

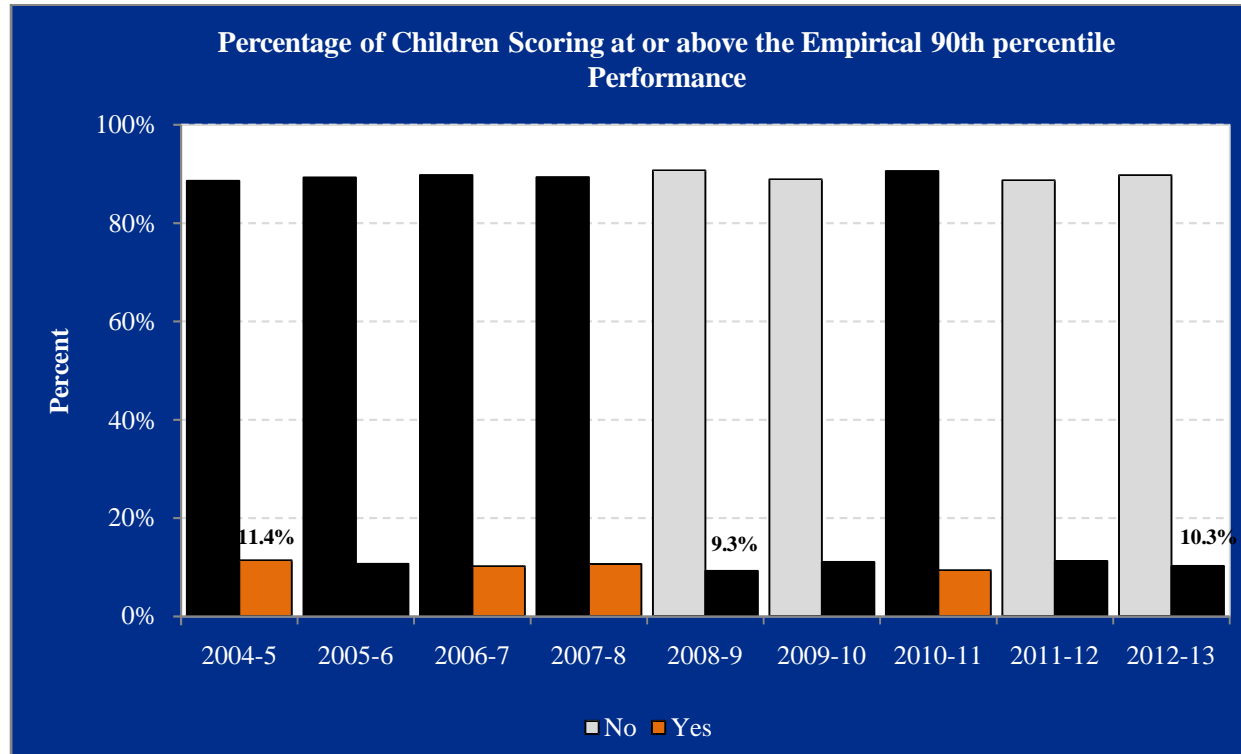
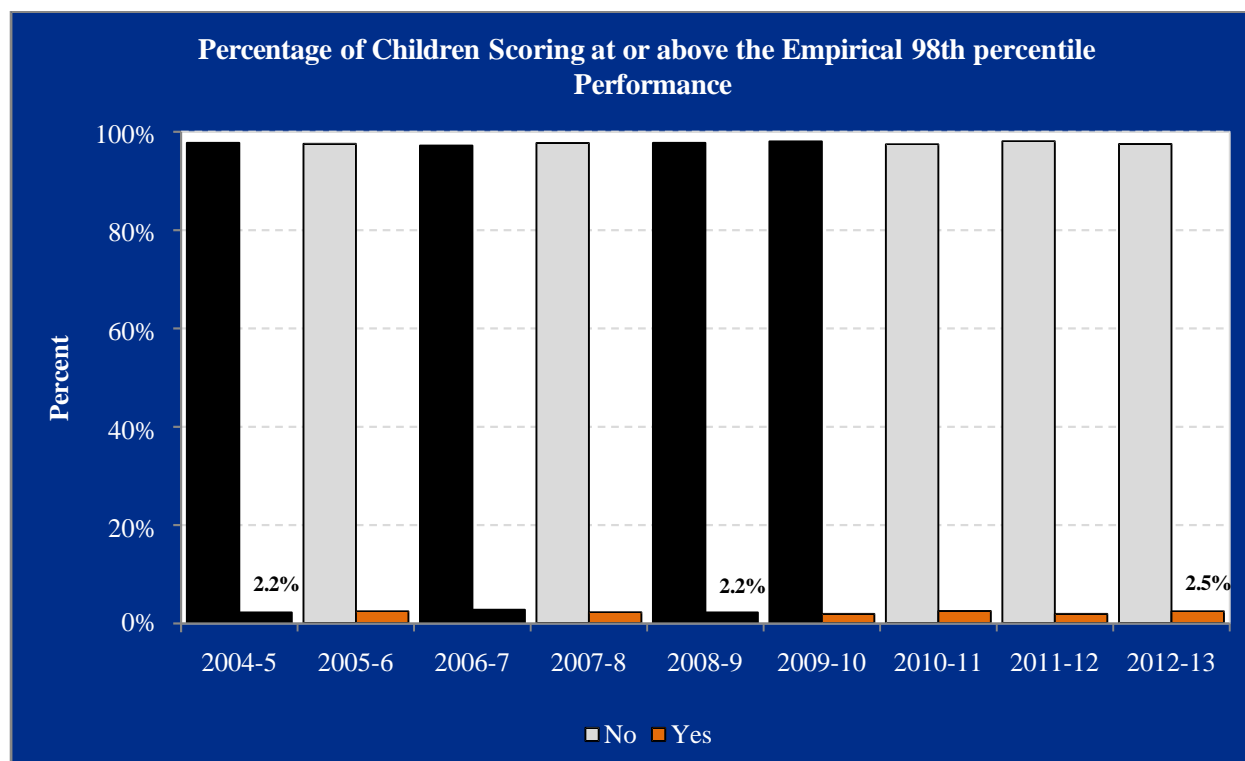


Figure 19. Percentage of Children At or Above the 90<sup>th</sup> Empirical Percentile – Performance





*Figure 20. Percentage of Children At or Above the 98<sup>th</sup> Empirical Percentile – Performance*





## Summary

The purpose of this research report was to evaluate recent trends in test-taker characteristics and mean ERB adapted WPPSI-III scores across a nine-year time frame. The general aims of this study were: to evaluate if there were certain trends in the demographic composition of the ERB samples across time; if mean scores on the WPPSI-III composite scores have significantly increased over time; and if the proportion of children scoring at the 90<sup>th</sup> and 98<sup>th</sup> percentiles on the assessment have increased substantially across years. In addition, empirical percentile rankings for the ERB sample were calculated and evaluated in comparison to the nationally normed percentile ranks that are currently used on the assessment for reporting purposes to see if similar patterns emerged.

The driving force behind this study was two-fold. First, no research to date has evaluated the ERB adapted WPPSI-III across time to assess the demographic characteristics of the children taking the test. Secondly, due to an overwhelming public perception within schools and across the public domain that all children taking this specific assessment provided by ERB are scoring at the top end of the distribution and that scale scores are increasing dramatically over recent years, this study attempted to gain valuable insight into whether this was indeed happening.

Several issues are worthy of discussion based on the results presented here. First, although the demographic composition of children taking the ERB adapted WPPSI-III appears to be relatively consistent across years in relation to age and gender, there is some evidence to suggest that children taking the WPPSI-III are about one to two months older in recent administrations. Whether the small magnitude in age difference is impacting the mean scores is unknown, but it could be posited that with such large developmental gains throughout the early developmental years of children, this small age difference may have a subtle impact on resulting scale scores.

Another trend that was apparent from results presented here was that there was a change in the demographic composition related to *when* children are testing. Over the course of nine years, there have been significantly more Pre-Kindergarten children testing over time. This finding coincides with the mean increase in the children testing, as Pre-Kindergarten children are generally older than those in Nursery school. This change concerning when children are testing could be a result of policy changes that have been implemented in the admission process for the



schools that use the ERB adapted WPPSI-III or it may be that more children are waiting slightly longer to take the test. Future research should evaluate the characteristics of schools using the WPPSI-III as an admission tool to understand if policies are impacting the demographic composition.

When looking at gender differences, this study highlighted that females score better overall than males. This finding is not particularly new as past research with clinical samples has indicated that, at younger ages, females tend to outperform males on overall scores (see Sattler & Dumont, 2004; Fisher, 2013). As research has also indicated that differences are found between males and females at higher and lower functioning levels, more research is warranted that disaggregates performance on the ERB adapted WPPSI-III scores by ability to assess if the same characteristics found in clinical samples can be generalized to this more able sample.

Evaluation of mean WPPSI-III scores across the aggregate sample indicated several key findings. First, it was found that, although there have been mean score increases in the full scale and performance scores, these increases have been very gradual over time — not limited to one or two administration periods. Second, it was found that children's performance on the verbal composite score has remained quite stable across administrations. Very few fluctuations were seen in the verbal scores, and it was even found that these scores have decreased slightly over the course of several recent years. This finding implicitly highlights that the variation that is seen in the full scale scores over time is a direct artifact of the variation found in children's performance scores, as the full scale score is simply the summation of the verbal and performance composites.

Evaluation of the magnitude of mean score changes across years indicated that there were no practically significant mean differences from one administration to the next. This finding is extremely important at a conceptual level in light of the heated debate about the increased availability of test preparatory materials and how this may be impacting overall scores. If one assumes that the majority of children are excessively utilizing these test preparatory tools, mean scores would potentially reflect this and would be increasing at alarming rates—especially in the most recent administration periods when these tools have become more readily available. Trends in mean scores did not support any type of drastic increase in recent administrations.

Another major concern for users and the general public is in relation to an increased fear that there are significantly more children scoring at the upper percentile ranks on the ERB adapted WPPSI-III. To evaluate if this was indeed the case, the proportion of children scoring at the 90<sup>th</sup> and 98<sup>th</sup> percentile ranks were evaluated over time. One can see from results presented here that over the course of nine years, although there have been marginal increases in the percentage of children scoring at these percentiles in general, over the most recent administrations (i.e., 2010-11 to 2012-13), there's been less than a 1% increase in the number of children scoring at or above the full scale score 90<sup>th</sup> percentile, about a 2 *decrease* in the proportion of children scoring at the verbal 90<sup>th</sup> percentile, and about a 5% increase in the proportion of children scoring at the performance score 90<sup>th</sup> percentile.

What is worth noting is the alarming number of children who score at these percentile ranks, *in general*. Across all administrations, approximately 62% to 70% of the children are reaching the 90<sup>th</sup> percentile and approximately 18% to 29% of the children are scoring at the 98<sup>th</sup> percentile. Results indicate that, although there are a large number of children are scoring at the 90<sup>th</sup> and 98<sup>th</sup> percentiles, this is not a new trend and the number of children scoring at these percentiles, as mentioned above, has not changed severely.

Reporting the percentage of children falling at or above these percentile ranks may be more misleading than helpful. This is largely because the ERB sample is much different from the standardization sample that was used to create these assigned percentile ranks. As was seen above, the children taking the ERB adapted WPPSI-III for admission are scoring, on average, 10 points above the national standardization sample, which implicitly suggests that the samples being evaluated are not normally distributed and are not directly comparable.

To illustrate the extreme inflation that can be observed in percentile rankings if these percentile ranks are used on samples that are inherently different, empirical percentile ranks for each ERB administration samples were calculated and evaluated. Results indicated that, when comparing the ERB samples to each other using a 'local norm' percentile ranking, less than 11% of the sample across administrations score at the 90<sup>th</sup> percentile, and roughly 2% of the sample score at



the 98<sup>th</sup> percentile within each administration. There were essentially no changes in the proportion of children scoring at or above these percentile ranks across the nine-year time frame. This illustration ultimately sheds light on a more pressing issue — one that questions the appropriateness of using the nationally normed WPPSI-III percentile rankings in the ERB population, which is a much more able sample than the national standardization sample.

Overall, evaluation of the ERB adapted WPPSI-III data used in this study revealed some important trends. The demographic composition of the sample across time was evaluated and important trends were highlighted. In addition, WPPSI-III score distributions and mean composite scores were evaluated, and each were found to be relatively consistent across the years. Although small gains have been made, nothing disparate was observed between any of the consecutive administrations. The same was found when evaluating the percentage of children scoring at the upper percentile ranks. Although the use of nationally normed percentiles is inflating the number of children scoring at these percentiles, there has been very little change in the number of children at or above these percentiles across recent years. The use of empirical percentile ranks for this sample was also highlighted, and may be a more discrete way to assess children's relative percentile ranks in this population. No substantial trends were observed when empirical percentile ranks were evaluated across time; however, the percentage of children scoring at these percentile ranks generally decreased from the 2004-5 to the 2012-13 year.

### **Limitations & Future Directions**

Although this report presents several key findings in relation to trends in test-taker characteristics and mean scores across time, it was intended to be a high-level evaluation of the trends associated with the demographic composition and mean scores across time. Further research should evaluate the effects of gender and age, and the interaction of the two on the WPPSI-III composite scores. Additionally, research should thoroughly evaluate the predictive validity of the WPPSI-III in the context of the admission process. Understanding if children are being correctly placed into programs — and are successful in these programs — is of major importance and should be of primary concern.

In addition, although the use of local norms may help assuage users' concerns about inflated percentile ranks, further research is definitely warranted. It may be that aggregating data



across multiple years (i.e., three years rolling norms) versus using one year of data provides a more stable and accurate way to rank performance via percentile ranks.

Lastly, although results from this study may help to alleviate concerns about the potential excessive use of test preparatory materials in preparation for this assessment, it by no means directly assessed this problem. Unfortunately, assessing if children are being influenced by tutoring or coaching is very difficult to do, as this type of information is not reported by schools or families and these services, for the most part, are used surreptitiously. More research is needed using data that can isolate whether or not children have access to these types of materials, and whether the use of these materials impact scores.



## References

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Second Edition. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Fisher, Barbara (2013). *What you think ADD/ADHD is, it isn't. Symptoms and Neuropsychological Testing Through Time*. Boca Raton, FL: Taylor & Francis Group, LLC.
- Sattler, J.M., & Dumont, R. (2004). *Assessment of children: WISC-IV and WPPSI-III supplement*. San Diego: Jerome M. Sattler Publisher, Inc.
- Wechsler, D. (2012). *The Wechsler Preschool and Primary Scale of Intelligence™ – Fourth Edition (WPPSI™ – IV)*. NCS Pearson, Inc.
- Wechsler, D. (2002). *The Wechsler Preschool and Primary Scale of Intelligence™, Third Edition (WPPSI™-III)*. San Antonio, TX: The Psychological Corporation
- What to expect on the ECAA One to One Candidates for Pre-K to Grade 4*, Educational Records Bureau, New York, NY, 2013.

## Appendix A

### Descriptive Statistics of the ERB Adapted WPPSI-III Scores Disaggregated by Gender Across Administrations

Table A.1. Mean Full Scale Scores by Gender Across Administrations

Administration	Full Scale Composite Score					
	Male			Female		
	<i>N</i>	Mean	SD	<i>N</i>	Mean	SD
2004 - 2005	1381	106.41	13.89	1292	108.90	13.03
2005 - 2006	1288	107.12	14.45	1352	110.31	13.22
2006 - 2007	1355	105.60	14.71	1365	108.78	14.37
2007 - 2008	1488	106.06	14.24	1598	108.53	13.64
2008 - 2009	1501	108.15	13.64	1540	110.86	12.62
2009 - 2010	1485	108.98	13.39	1380	111.26	12.92
2010 - 2011	1571	109.62	14.06	1548	113.30	13.37
2011 - 2012	1500	110.16	14.14	1554	112.84	13.67
2012 - 2013	1388	110.55	14.49	1328	113.34	13.81

Table A.2. Mean Verbal Scale Scores by Gender Across Administrations

Administration	Verbal					
	Male			Female		
	<i>N</i>	Mean	<i>N</i>	Mean	<i>N</i>	Mean
2004 - 2005	1381	56.31	8.85	1292	57.25	8.17
2005 - 2006	1288	56.46	8.89	1352	57.78	8.44
2006 - 2007	1355	55.77	9.24	1365	57.09	8.98
2007 - 2008	1488	55.73	8.54	1598	56.98	8.32
2008 - 2009	1501	56.73	8.30	1540	58.15	7.75
2009 - 2010	1485	56.99	8.08	1380	58.28	7.75
2010 - 2011	1571	57.14	8.19	1548	58.59	7.88
2011 - 2012	1500	56.84	8.22	1554	58.33	7.86
2012 - 2013	1388	56.88	8.45	1328	58.18	8.13

Table A.3. Mean Performance Scale Scores by Gender Across Administrations

Administration	Performance					
	Male			Female		
	<i>N</i>	Mean	<i>N</i>	Mean	<i>N</i>	Mean
2004 - 2005	1381	50.09	7.47	1292	51.65	7.04
2005 - 2006	1288	50.67	7.80	1352	52.53	6.99
2006 - 2007	1355	49.83	7.97	1365	51.69	7.60
2007 - 2008	1488	50.33	8.01	1598	51.55	7.47
2008 - 2009	1501	51.42	7.63	1540	52.71	7.02
2009 - 2010	1485	51.99	7.65	1380	52.98	7.27
2010 - 2011	1571	52.48	8.08	1548	54.72	7.36
2011 - 2012	1500	53.32	8.16	1554	54.51	7.72
2012 - 2013	1388	53.67	8.22	1328	55.16	7.75



## Appendix B

### Frequency Distributions of the ERB Adapted WPPSI-III Full Scale Scores by Administration

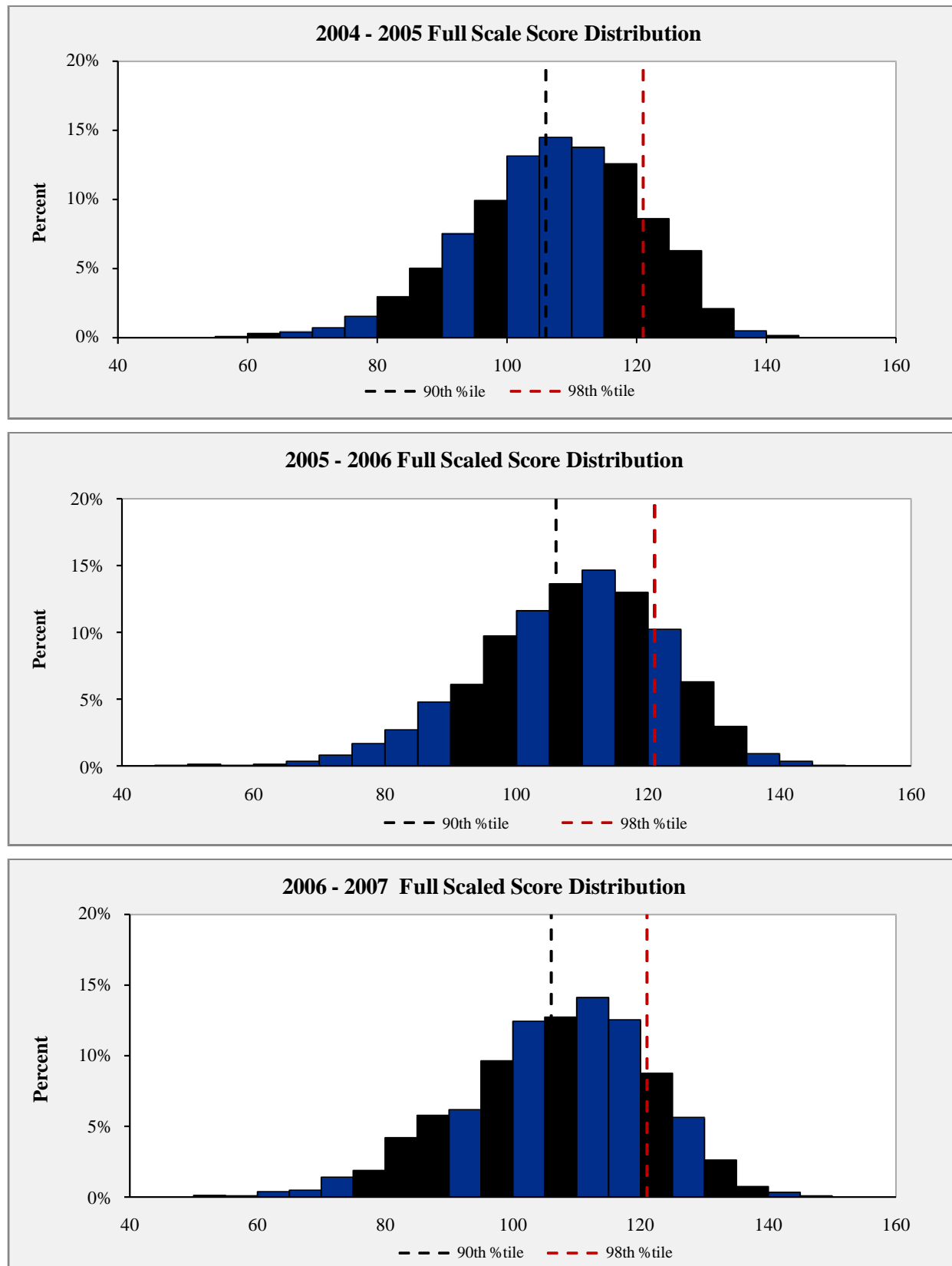


Figure B.1. WPPSI-III Full Scale Score Distributions: 2004-5 to 2006-7

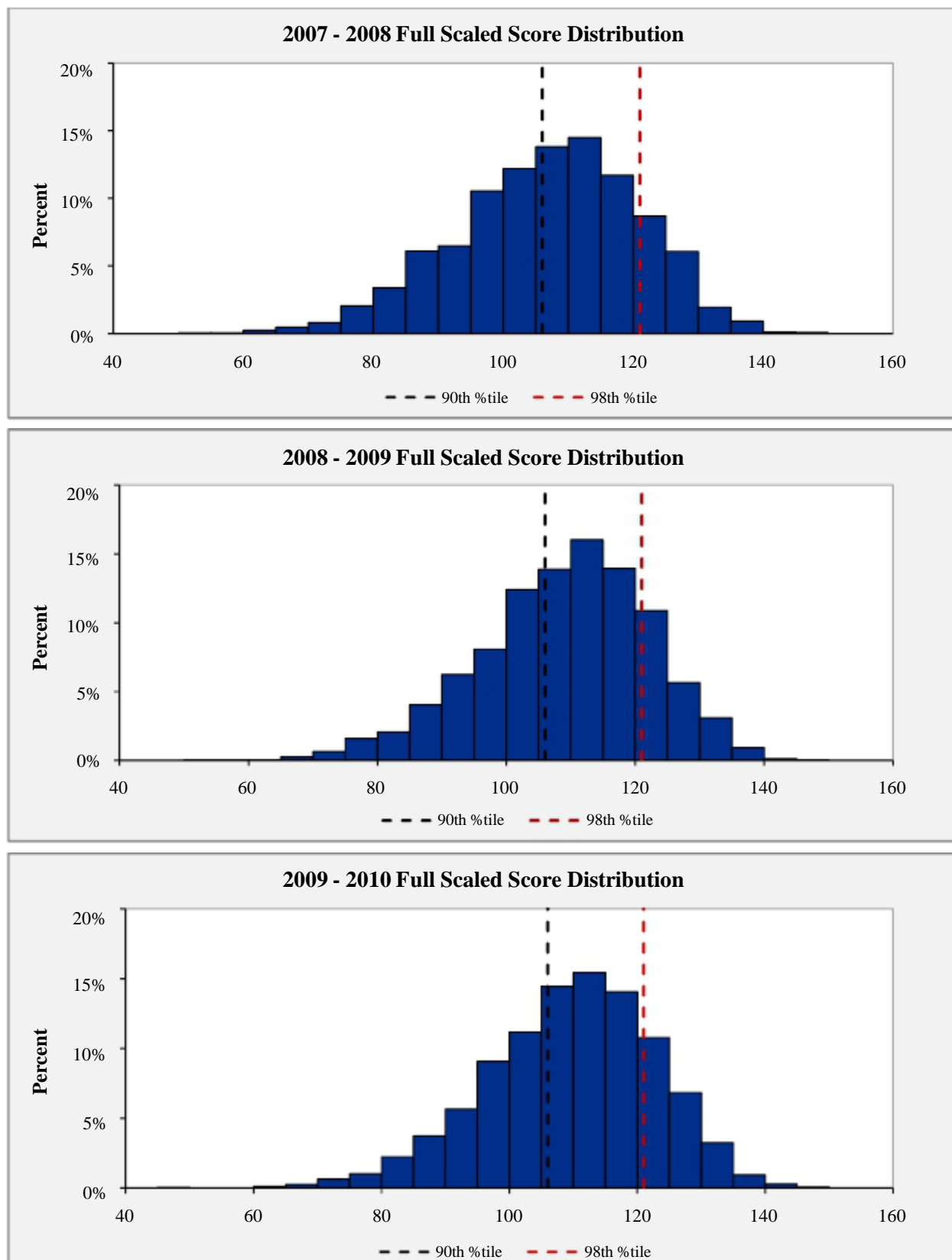


Figure B.2. WPPSI-III Full Scale Score Distributions: 2007-8 to 2009-10

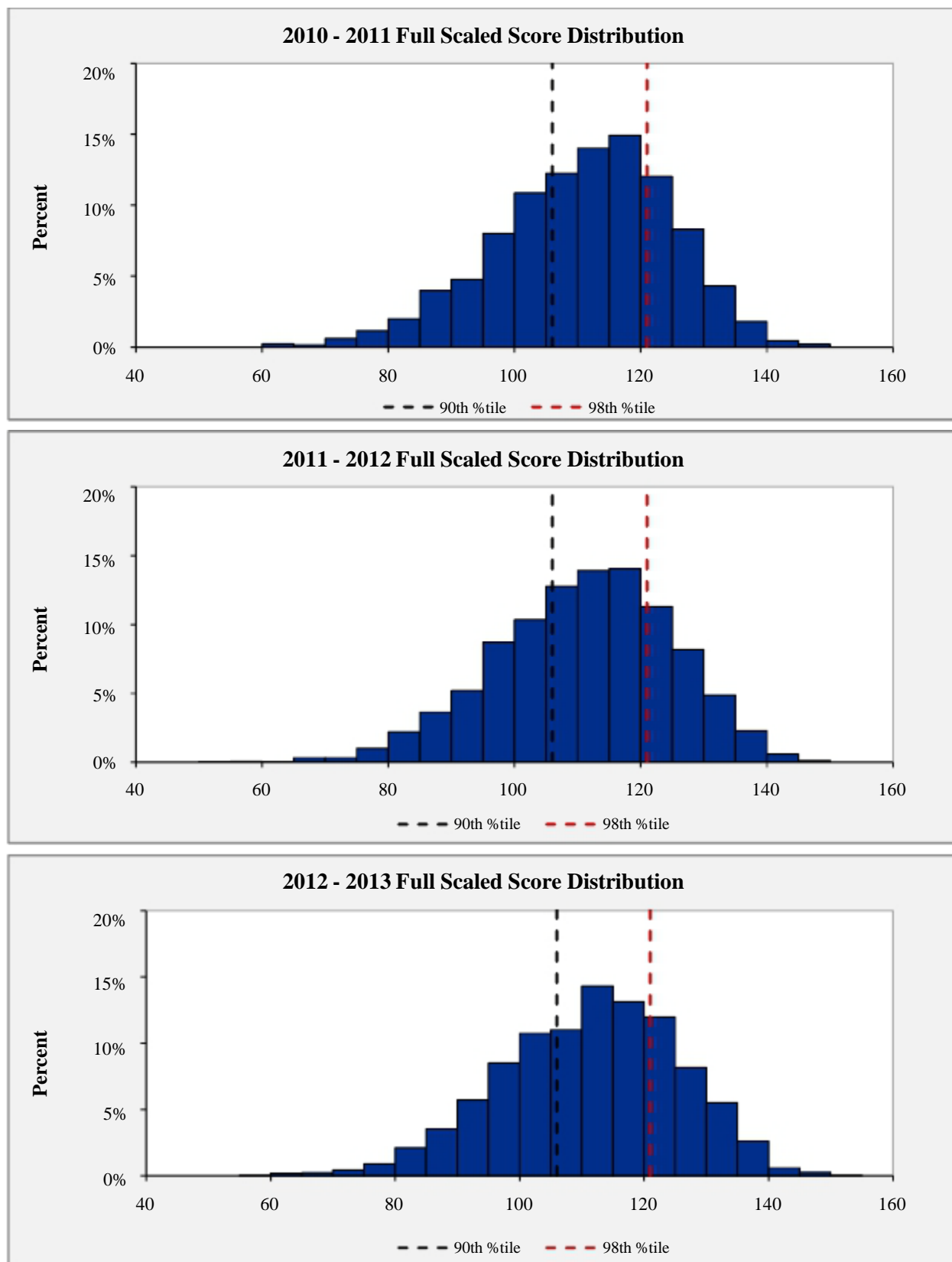


Figure B.3. WPPSI-III Full Scale Score Distributions: 2010-11 to 2012-13



## Appendix C

### Frequency Distributions of the ERB Adapted WPPSI-III Verbal Scores by Administration

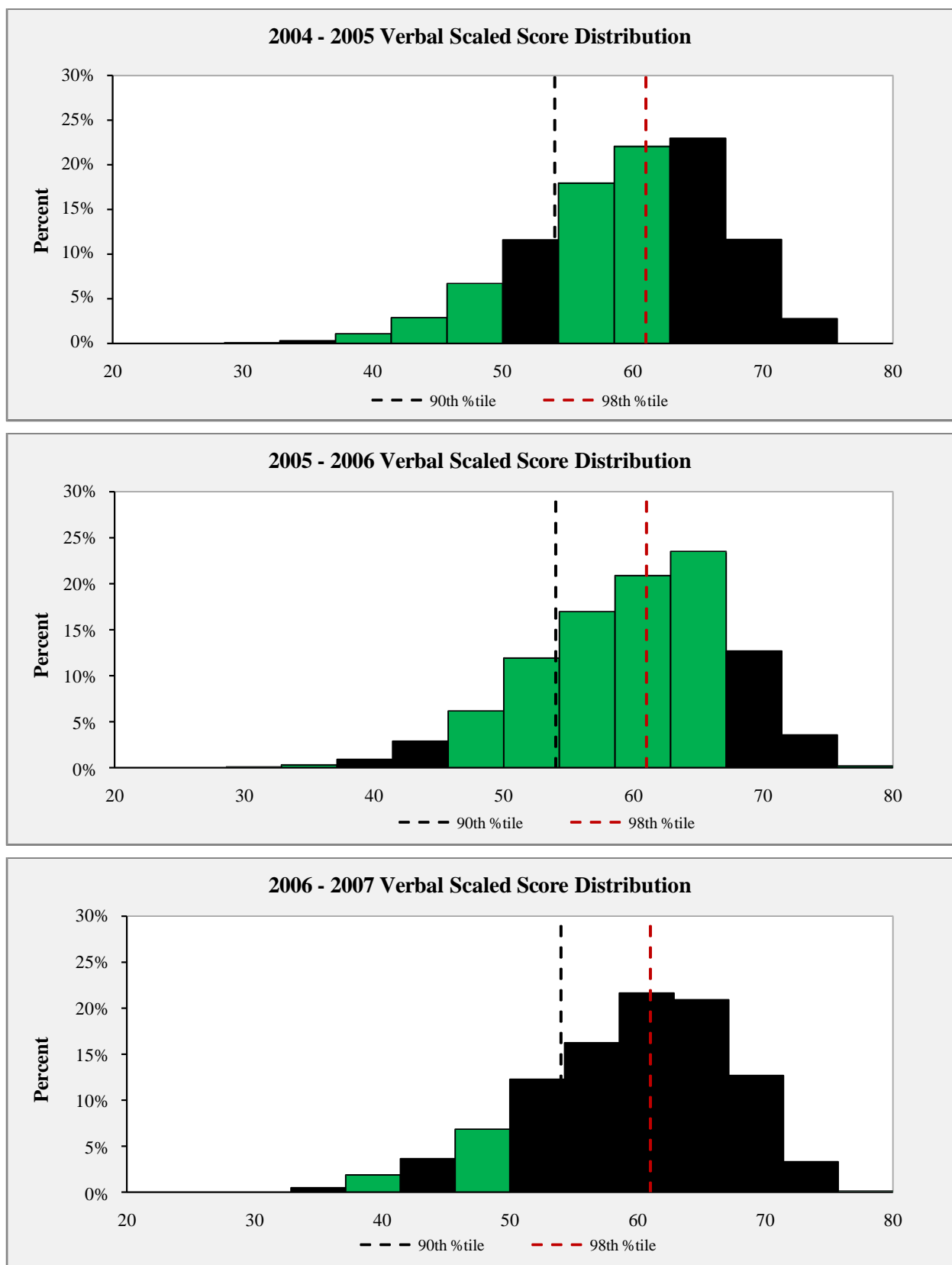


Figure C.1. WPPSI-III Verbal Score Distributions: 2004-5 to 2006-7

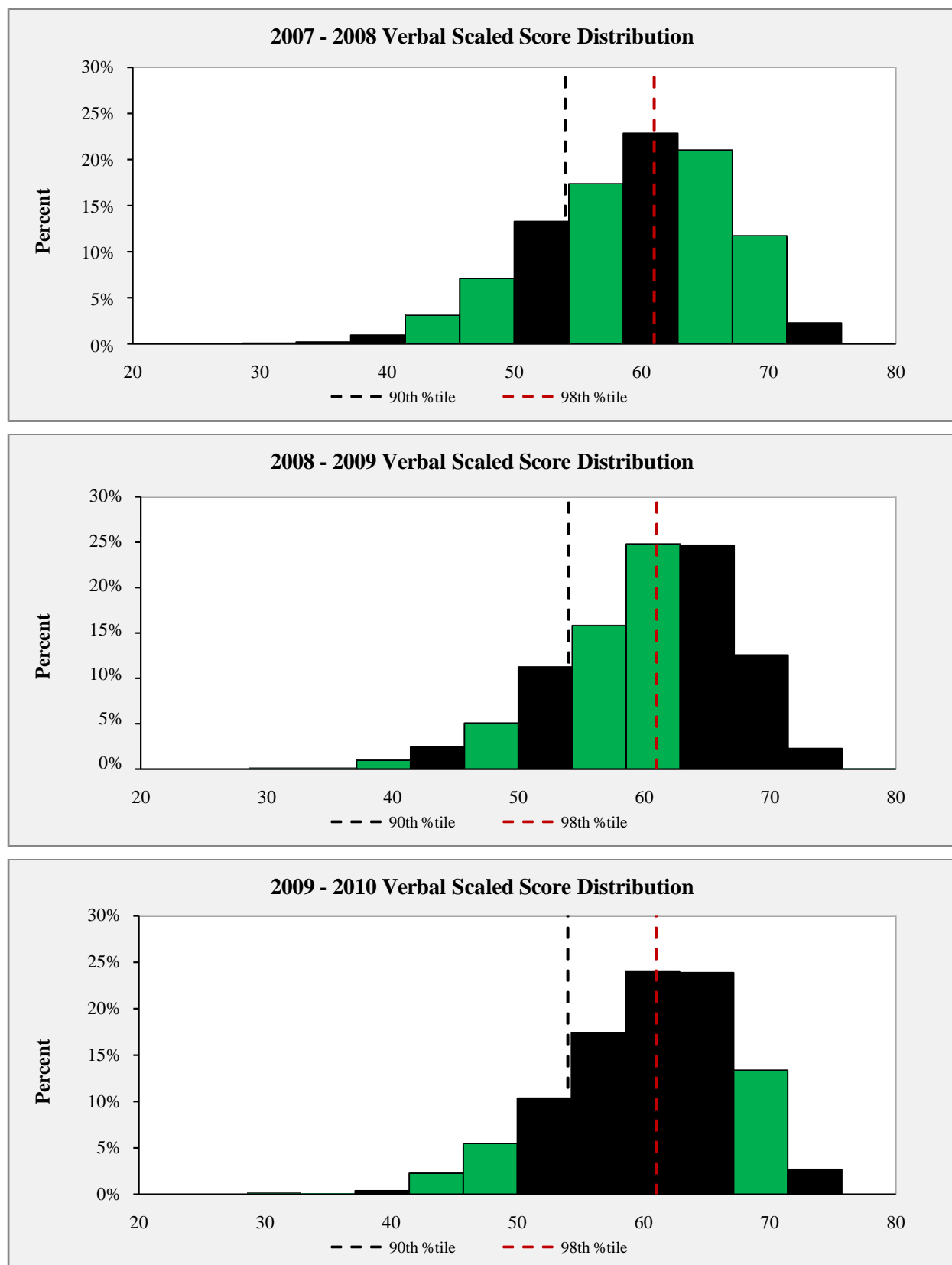


Figure C.2. WPPSI-III Verbal Score Distributions: 2007-8 to 2009-10

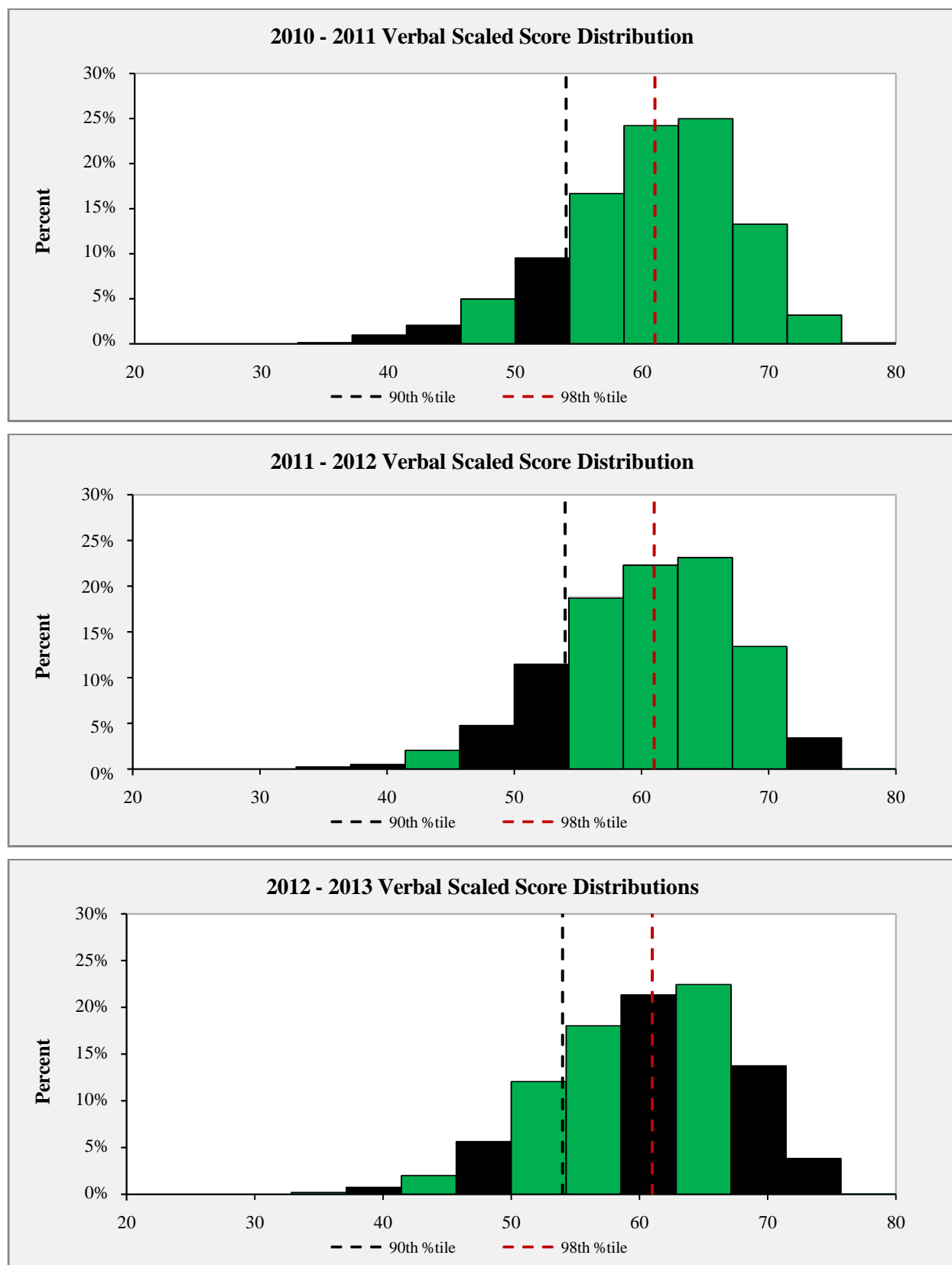


Figure C.3. WPPSI-III Verbal Score Distributions: 2010-11 to 2012-13



## Appendix D

### Frequency Distributions of the ERB Adapted WPPSI-III Performance Scores by Administration

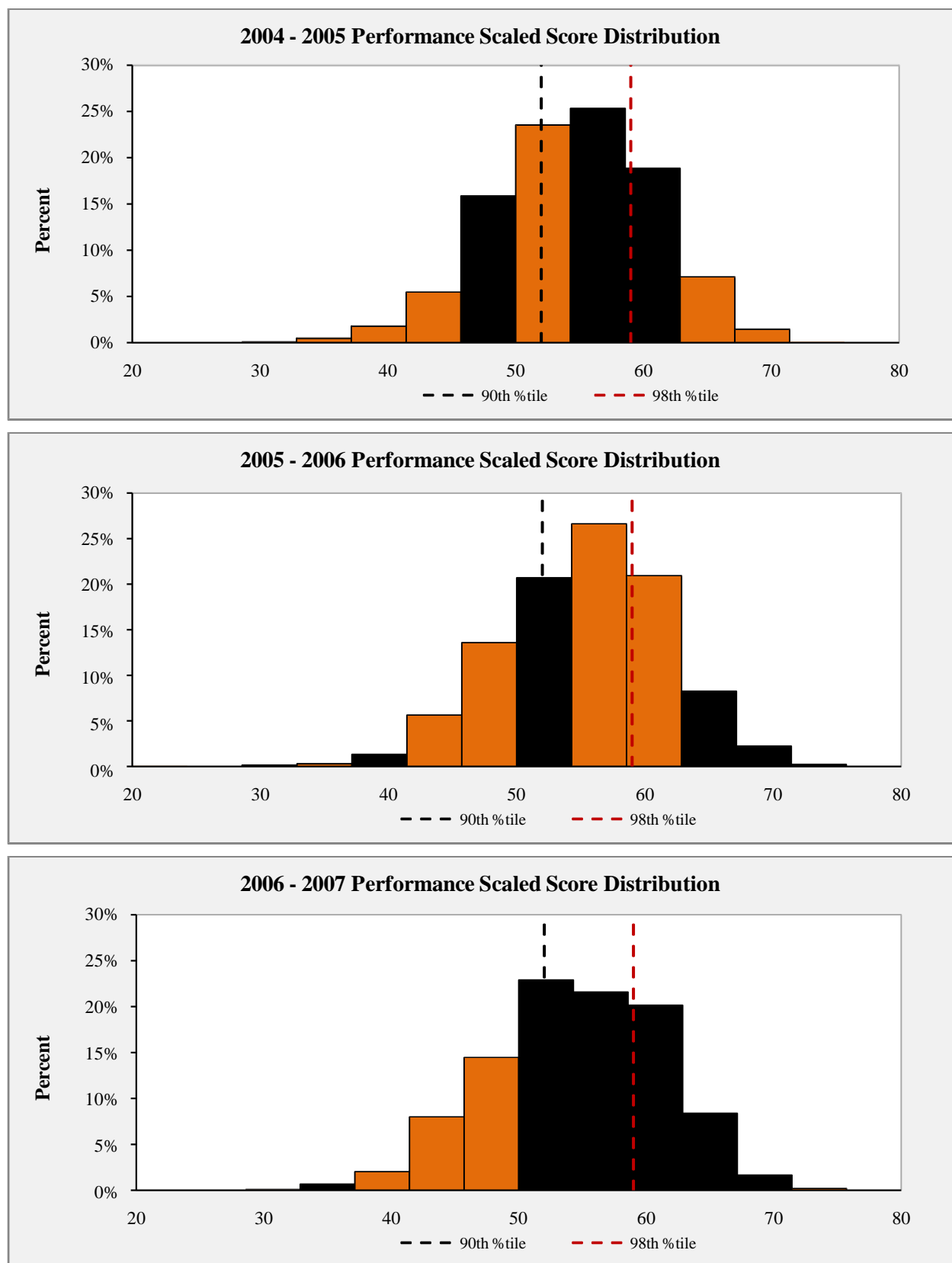


Figure D.1. WPPSI-III Performance Score Distributions: 2004-5 to 2006-7

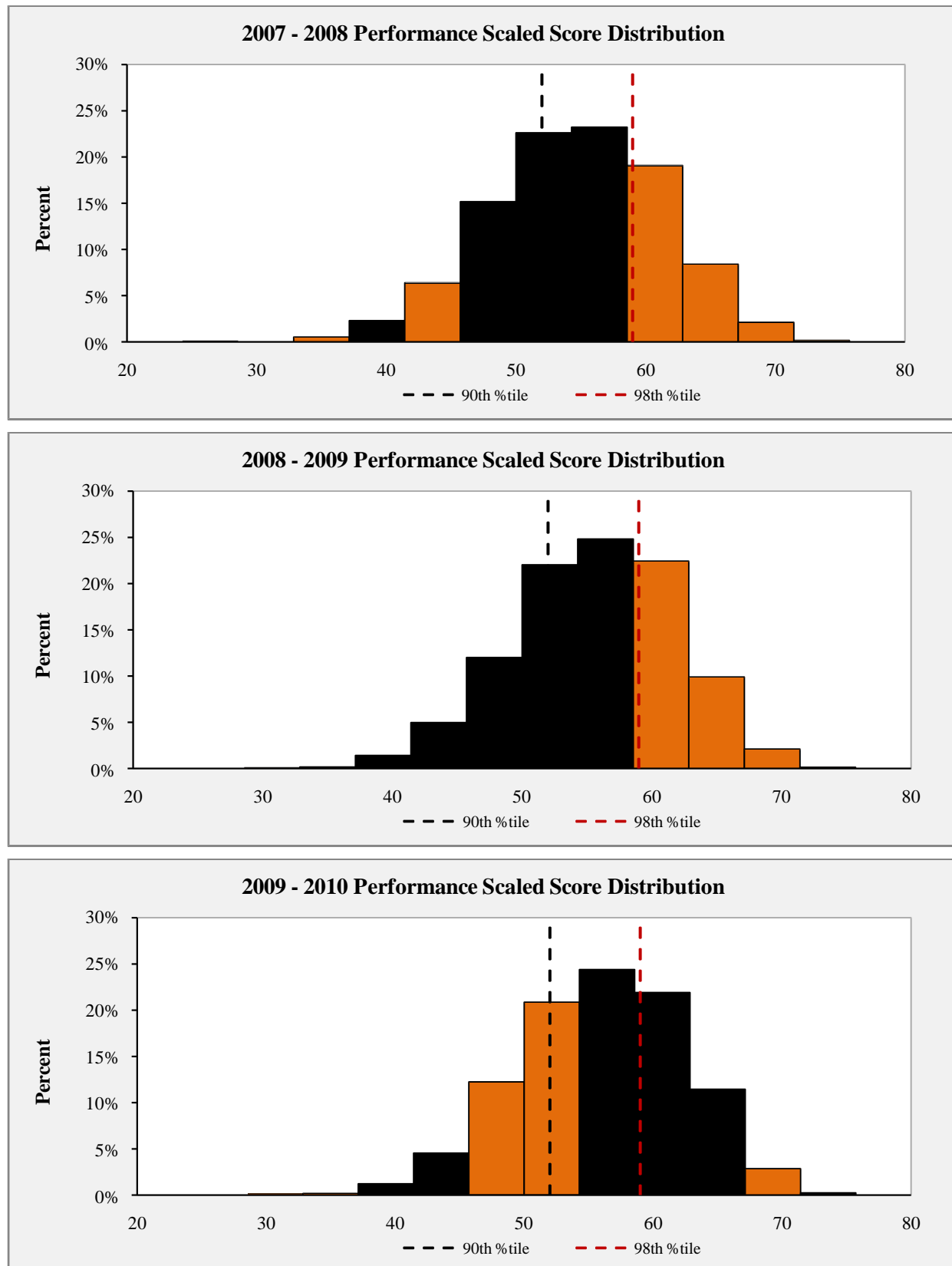


Figure D.2. WPPSI-III Performance Score Distributions: 2007-8 to 2009-10

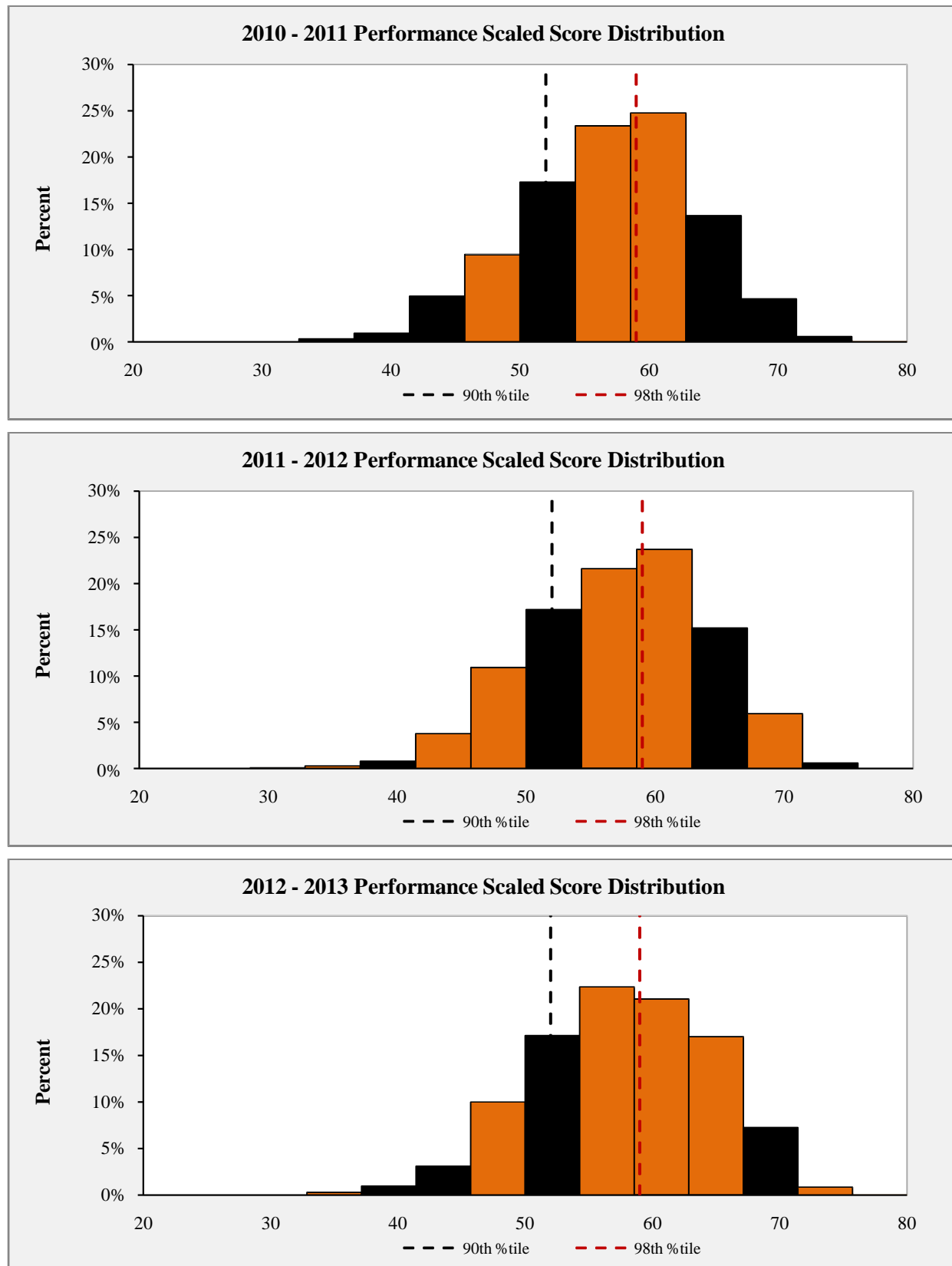


Figure D.3. WPPSI-III Performance Score Distributions: 2010-11 to 2012-13

