

# Appendix

## Appendix A1 Study characteristics: Kemple, Herlihy, & Smith, 2005 (quasi-experimental design)

Characteristic	Description
<b>Study citation</b>	Kemple, J. J., Herlihy, C. M., & Smith, T.J. (2005). <i>Making progress toward graduation: Evidence from the Talent Development High School model</i> . New York: MDRC.
<b>Participants</b>	<p>The main analysis sample included first-time ninth-grade students<sup>1</sup> from five high schools that began implementing <i>Talent Development High Schools</i> between 1999 and 2001 and six matched comparison high schools.<sup>2</sup> Between two and four comparison schools were matched to each of the five intervention schools based on the racial/ethnic composition and promotion rates of the schools' ninth-grade students (Kemple &amp; Herlihy, 2004). A comparison school could be matched to multiple <i>Talent Development High Schools</i>. The study compared the outcomes of ninth graders who entered <i>Talent Development High Schools</i> in the three years immediately after the program was implemented with those of ninth graders from these schools in the three years just before program implementation and with the outcome differences over the same time period for the matched comparison schools.<sup>3</sup></p> <p>Many students selected for <i>Talent Development High Schools</i> had low test scores and were overage for their grade. More than three-quarters were African-American and about one in six were Hispanic. Poor attendance was common, with two-thirds missing at least 20% of scheduled school days during their ninth-grade year. In addition, many did not make regular progress toward graduation, with just half promoted to tenth grade at the end of their ninth-grade year. Students in the matched comparison schools were generally similar to <i>Talent Development High Schools</i> students on these characteristics (Kemple &amp; Herlihy, 2004).</p> <p>The study examined three cohorts of students. Cohort 1 included students in the intervention and matched comparison schools who enrolled in the ninth grade during the first year of <i>Talent Development High Schools</i> implementation at the intervention schools. Similarly, Cohort 2 and Cohort 3 included students who were enrolled in the ninth grade during the second and third years of implementation, respectively. Given the fixed period for data collection, later cohorts had shorter follow-up periods. To ensure both an adequate follow-up and an adequate sample size for assessing program effectiveness, the WWC used second-year results based on Cohorts 1 and 2 to rate the effectiveness of <i>Talent Development High Schools</i>. Longer-term results based only on Cohort 1 and shorter-term results based on all three cohorts are reported in Appendix A4.</p>
<b>Setting</b>	The impact study was conducted in 11 nonselective public high schools in Philadelphia.
<b>Intervention</b>	<p>The Philadelphia public school district implemented the <i>Talent Development High Schools</i> model in seven high schools. The district began to roll out the program in 1998, with one or two high schools launching <i>Talent Development High Schools</i> each year over a five-year period. School administrators volunteered their schools as candidates for implementing the new program. To allow for adequate follow-up, the impact study excluded the two Philadelphia high schools that implemented <i>Talent Development High Schools</i> last.</p> <p>All the Philadelphia <i>Talent Development High Schools</i> created ninth-grade academies on a separate floor or wing of the building, which were taught by teams of four to five teachers. Each school introduced block scheduling with 80- to 90-minute class sessions, introducing “double dose” math and English courses for ninth and tenth graders. These double sections of English and math allowed students to both prepare for and take college preparatory classes over the course of one academic year. Six of the seven schools offered “Twilight School” for new or repeating ninth graders with serious attendance or discipline problems.<sup>4</sup></p> <p>The model for students in grades 10 through 12 centered around career academies, in which students were divided into smaller “learning communities” around a broad career interest and the curriculum was organized around a career theme. Many Philadelphia high schools already had career academies before <i>Talent Development High Schools</i> was implemented, including many non-<i>Talent Development</i> schools. The study authors concluded that “(i)t is likely, therefore, that the upper-grade experience of students in <i>Talent Development</i> schools did not greatly differ from that of students in non-<i>Talent Development</i> schools” (Kemple, Herlihy, &amp; Smith, 2005, p. 27).</p> <p>The study authors reported some variation in how the program was implemented across schools (Kemple, Herlihy, &amp; Smith, 2005). In particular, they noted considerable variation across the intervention schools in the amount of technical assistance and support schools received from the intervention developer, as well as the amount of intervention-specific training school staff received.</p>

(continued)

## Appendix A1 Study characteristics: Kemple, Herlihy, & Smith, 2005 (quasi-experimental design) *(continued)*

Characteristic	Description
<b>Comparison</b>	Matched comparison schools were nonselective Philadelphia high schools that did not implement <i>Talent Development High Schools</i> . The authors compared the intervention group both with students in the comparison schools and with students who attended the intervention schools prior to the implementation of <i>Talent Development High Schools</i> .
<b>Primary outcomes and measurement</b>	<p>Two relevant outcomes are included in this review: total credits earned and enrollment in the tenth grade by the end of the second year of high school (see Appendix A2 for more detailed descriptions of these outcome measures).</p> <p>The study also examined <i>Talent Development High Schools</i>' effects on attendance and student achievement. These outcomes are not included in this report because they do not fall within the three domains (staying in school, progressing in school, and completing school) examined by the WWC's review of dropout prevention interventions. Effects on the percentage of students who exited the school district were also estimated. However, the WWC had concerns about the validity of this measure and did not include it in the review.<sup>5</sup> The study also examined the effects of <i>Talent Development High Schools</i> on graduation on the two earliest implementing schools. Since these results are only available for a small subset of the full research sample, they are not considered for the effectiveness rating and improvement index.</p>
<b>Teacher training</b>	Teachers at <i>Talent Development High Schools</i> were regular teachers employed by the Philadelphia Public Schools. "Curriculum coaches" who had been trained by the intervention developer provided on-site technical assistance with implementing the <i>Talent Development High Schools</i> model. The developer also provided summer training institutes for staff.

1. The authors also examined the effects of *Talent Development High Schools* on students repeating the ninth grade. However, this review focuses only on the effects of the intervention on first-time ninth graders.
2. The sample was restricted to students who attempted at least one course credit during the ninth grade and excluded special education students and English language learners. The authors did not report the number of students in the sample.
3. To take into account the potential influence of changes in the composition of students attending the study schools, the authors used ordinary least squares regression to control for a set of student background characteristics (race, seventh-grade reading and math test scores, and whether the student had repeated a grade) when estimating impacts.
4. Reports from the study do not indicate whether the one school that did not implement the "Twilight School" was among the five schools included in the impact analysis.
5. In particular, the measure did not appear to adequately capture the full extent of dropping out because only about 2% of students were categorized as having dropped out during the first two years of high school.

## Appendix A2 Outcome measures in the progressing in school domain

Outcome measure	Description
<b>Total credits earned by end of second year of high school</b>	This measure represents the cumulative number of course credits earned over the first two years of high school. These data were collected from individual students' school records obtained from the district.
<b>Enrolled in tenth grade by end of second year of high school</b>	This measure represents the percentage of students who were enrolled in the tenth grade by the end of the second year of high school. These data were collected from individual students' school records obtained from the district.

## Appendix A3 Summary of study findings included in the rating for the progressing in school domain<sup>1</sup>

Outcome measure	Study sample	Sample size (schools) <sup>2</sup>	Authors' findings from the study			WWC calculations		
			Mean outcome		Mean difference <sup>5</sup> (Talent Development High Schools – comparison)	Effect size <sup>6</sup>	Statistical significance <sup>7</sup> (at $\alpha = 0.05$ )	Improvement index <sup>8</sup>
			Talent Development High Schools group <sup>3</sup>	Comparison group <sup>4</sup>				
<b>Kempe, Herlihy, &amp; Smith, 2005 (quasi-experimental design)<sup>9</sup></b>								
Total credits earned by end of second year	Cohorts 1 and 2	11	9.5	8.6	0.9	0.16	Statistically significant	+6
Enrolled in tenth grade by end of second year (%)	Cohorts 1 and 2	11	68	60	8	0.21	Statistically significant	+8
<b>Domain average<sup>10</sup> for progressing in school across all studies</b>						0.18	Statistically significant	+7

1. This appendix reports follow-up findings considered for the effectiveness rating and the improvement index. Kempe, Herlihy, & Smith (2005) examined *Talent Development High Schools'* effects on three cohorts of entering ninth-grade students. Given the fixed period for data collection, the follow-up period varied by cohort. Cohort 1 had three years of follow-up, while Cohort 2 had two years, and Cohort 3 had only one year. To balance the benefits of assessing program effectiveness using a longer follow-up period with the benefits of assessing effectiveness using results from multiple cohorts of students, the WWC used results for Cohorts 1 and 2 measured at the end of the second year of high school. Follow-up findings from the end of the first year of high school (available for all three cohorts) and the third year of high school (available for Cohort 1 only) are not included in these ratings, but are reported in Appendix A4.
2. Kempe, Herlihy, & Smith (2005) used individual student data, but did not report the number of students in the sample.
3. The WWC calculated the cross-cohort average for each measure as the simple average of the mean outcomes for each cohort included in the measure.
4. Kempe, Herlihy, & Smith (2005) did not report these adjusted comparison group means and instead reported baseline to follow-up changes for both the intervention and comparison groups. The WWC generated the adjusted comparison group means reported here using the following transformation: adjusted comparison group mean = follow-up comparison group mean + (baseline intervention group mean – baseline comparison group). Stated differently, the adjusted comparison group mean equals the follow-up intervention group mean minus the estimated impact because, under the comparative interrupted time-series technique used in Kempe, Herlihy, & Smith (2005), impacts are calculated as follows: impact = (follow-up intervention group mean – baseline intervention group mean) – (follow-up comparison group mean – baseline comparison group mean).
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
6. The effect size for the “total credits earned” outcome was reported by the study authors. The study authors calculated the effect size by dividing the impact at follow-up by the standard deviation of the outcome for all ninth-grade students in the Philadelphia school district’s nonselective, comprehensive high schools from school years 1996/97 through 1998/99. The effect size for the dichotomous variable “enrolled in the tenth grade by the end of the second year” was computed using the Cox Index. For an explanation of the effect size calculations, see [Technical Details of WWC-Conducted Computations](#).
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
9. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Kempe, Herlihy, & Smith (2005), the study authors made appropriate corrections for clustering, so no additional corrections for clustering were necessary. The WWC corrected statistical significance levels for multiple comparisons.
10. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

## Appendix A4 Summary of shorter-term and longer-term findings for the progressing in school domain<sup>1</sup>

Outcome measure	Study sample	Sample size (schools) <sup>2</sup>	Authors' findings from the study			WWC calculations		
			Mean outcome		Mean difference <sup>5</sup> (Talent Development High Schools – comparison)	Effect size <sup>6</sup>	Statistical significance <sup>7</sup> (at $\alpha = 0.05$ )	Improvement index <sup>8</sup>
			Talent Development High Schools group <sup>3</sup>	Comparison group <sup>4</sup>				
<b>Kempe, Herlihy, &amp; Smith, 2005 (quasi-experimental design)<sup>9</sup></b>								
Total credits earned by end of first year	Cohorts 1, 2, and 3	11	5.2	4.5	0.7	0.25	Statistically significant	+10
Total credits earned by end of third year	Cohort 1	11	13.2	12.3	0.9	0.12	ns	+5
Enrolled in eleventh grade by end of third year	Cohort 1	11	53	47	6	0.16	ns	+6

1. This appendix presents follow-up findings at the end of the first and third years of high school for the measures that fall in the progressing in school domain. The third-year findings were based only on students who began the ninth-grade during the first year of *Talent Development* implementation (cohort 1). The first-year findings were based on all three cohorts. The second-year findings used for effectiveness rating purposes are presented in Appendix A3.
2. Kempe, Herlihy, & Smith (2005) used individual student data, but did not report the number of students in the sample.
3. The WWC calculated the cross-cohort average for the “total credits earned by end of first year” outcome as the simple average of the mean outcomes for each cohort included in the measure.
4. Kempe, Herlihy, & Smith (2005) did not report these adjusted comparison group means and instead reported baseline to follow-up changes for both the intervention and comparison groups. The WWC generated the adjusted comparison group means reported here using the following transformation: adjusted comparison group mean = follow-up comparison group mean + (baseline intervention group mean – baseline comparison group). Stated differently, the adjusted comparison group mean equals the follow-up intervention group mean minus the estimated impact, since, under the comparative interrupted time-series technique used in Kempe, Herlihy, & Smith (2005), impacts are calculated as follows: impact = (follow-up intervention group mean – baseline intervention group mean) – (follow-up comparison group mean – baseline comparison group mean).
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
6. The effect size for the “total credits earned” outcome was reported by the study authors. The study authors calculated the effect size by dividing the impact at follow-up by the standard deviation of the outcome for all ninth-grade students in the Philadelphia school district’s nonselective, comprehensive high schools from school years 1996/97 through 1998/99. The effect size for the dichotomous variable “enrolled in the tenth grade by the end of the second year” was computed using the Cox Index. For an explanation of the effect size calculations, see [Technical Details of WWC-Conducted Computations](#).
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
9. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools. For an explanation, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate the statistical significance. In the case of Kempe, Herlihy, & Smith (2005), the study authors made corrections for clustering and no additional corrections for clustering was necessary.

## Appendix A5 Talent Development High Schools rating for the progressing in school domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup>

For the outcome domain of progressing in school, the WWC rated *Talent Development High Schools* as having potentially positive effects. It did not meet the criteria for positive effects because there was only one study and that study did not meet WWC evidence standards for a strong design. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Talent Development High Schools* was assigned the highest applicable rating.

### Rating received

**Potentially positive effects:** Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

**Met.** One study of *Talent Development High Schools* demonstrated a statistically significant positive effect.

#### AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

**Met.** No studies found statistically significant or substantively important negative effects in this domain.

### Other ratings considered

**Positive effects:** Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

**Not met.** *Talent Development High Schools* had only one study meeting WWC evidence standards.

#### AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

**Met.** No studies found statistically significant or substantively important negative effects in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the [WWC Intervention Rating Scheme](#) for a complete description.

## Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence <sup>1</sup>
		Schools	Students	
Staying in school	0	0	0	na
Progressing in school	1	11	nr	Small
Completing school	0	0	0	na

na = not applicable/not studied

nr = not reported

1. A rating of “moderate to large” requires at least two studies and two schools across studies in one domain, and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.”