# What Works Clearinghouse™

ies INSTITUTE OF EDUCATION SCIENCES

Adolescent Literacy January 2013

# Talent Development Middle Grades Program

## Program Description[1]

The *Talent Development Middle Grades Program* is a comprehensive reform model that transforms the structure and curriculum of large urban middle schools with the aim of improving student achievement and raising teacher and student expectations. Key features of the *Talent Development Middle Grades Program* include small learning communities, an evidence-based curriculum tied to standards, the use of teacher teams, professional development and support for teachers, and school–family–community connections. This review focuses on the effects of the *Talent Development Middle Grades Program* on student reading outcomes. *Student Team Literature* is the literacy component of the *Talent Development Middle Grades Program*, and is therefore relevant to this review in the Adolescent Literacy topic area.

*Student Team Literature* is a reading and English language arts curriculum for middle school students that utilizes cooperative learning activities, high-interest reading materials, and explicit instruction to teach reading strategies, comprehension skills, and fluency in reading and writing. It offers integrated reading, writing, and language arts instruction and a writing process approach to language arts. As part of the program, students work in groups, and activities follow a regular cycle that involves teacher presentation, team practice, independent practice, peer preassessment, and individual assessments that form the basis for team scores. The cooperative learning teams are intended to engage students in academic interactions and create a motivating environment.

## Report Contents



## Research[2]

The What Works Clearinghouse (WWC) identified one study of the *Talent Development Middle Grades Program* that both falls within the scope of the Adolescent Literacy topic area and meets WWC evidence standards with reservations. The study included adolescent readers from grades 7 and 8 in 29 urban middle schools in the Northeast.

The WWC considers the extent of evidence for the *Talent Development Middle Grades Program* on the reading performance of adolescent readers to be small for one outcome domain: comprehension.[3]

## Effectiveness

The *Talent Development Middle Grades Program* was found to have potentially positive effects on comprehension for adolescent readers. Three other domains in this topic area are not covered in this intervention report. (See the Effectiveness Summary on p. 6 for further description of all domains.)

### Table 1. Summary of findings[4]

| Outcome domain | Rating of effectiveness | Improvement index *(percentile points)* | | Number of studies | Number of students | Extent of evidence |
|---|---|---|---|---|---|---|
| | | Average | Range | | | |
| **Comprehension** | Potentially positive | +3 | –3 to +7 | 1 | nr | Small |

nr = not reported

# Program Information

## Background

The *Talent Development Middle Grades Program* was developed at Johns Hopkins University by researchers, educators, and curriculum writers in collaboration with middle school practitioners. For more information, contact Kathy Nelson, Director of Implementation, Talent Development Middle Schools Program, Center for Social Organization of Schools, Johns Hopkins University. Address: 3003 N. Charles St., Suite 200, Baltimore, MD 21218. Email: knelson@csos.jhu.edu. Telephone: (410) 516-6431. Fax: (410) 516-8890.

## Program details

Key features of the *Talent Development Middle Grades Program* include small learning communities, an evidence-based curriculum tied to standards, the use of teacher teams, professional development and support for teachers, and school–family–community connections. Schools are systematically reorganized into small learning communities with interdisciplinary teacher teams. Each team shares the same students and has common planning time. In addition to *Student Team Literature*, curricula are also available in mathematics, science, and US history.

The *Talent Development Middle Grades Program* includes time for career exploration and courses in study skills and social skills. Teachers receive professional development on the use of the curriculum and accompanying instructional practice. Each school employs the services of curriculum coaches to provide ongoing teacher support. The program also provides catch-up opportunities during the school day for students who are struggling with mathematics or reading. The model has been implemented with culturally and linguistically diverse populations.

*Student Team Literature*[5] is part of the *Talent Development Middle Grades Program*. *Student Team Literature* has two principal elements: literature-related activities and direct instruction in reading comprehension. First, teachers introduce novels to the class and discuss relevant background knowledge, genre, and vocabulary. The program pairs *Talent Development Student Team Literature* discussion guides with high-interest trade books. Discussion guides are available for nearly 200 works, including over 60 award-winning novels and plays.

Teachers begin by introducing students to the text, and then organize students into learning teams of four or five for reading activities. They receive rewards for working well both as individuals and as group members. Students take part in the following literature-related activities:

- Partner Reading—students first read silently, then take turns reading orally with a partner.
- Treasure Hunts—questions guide student reading, requiring them to search and think to generate text-supported answers.
- Word Mastery—students practice saying new vocabulary words with their partners. They then use those words in story-related writing.
- Story-Retelling—students summarize stories in their own words.
- Story-Related Writing—students write in response to prompts about their reading.
- Extension Activities—students conduct cross-curricular research and complete activities related to fine arts, dramatics, and media as they explore themes in the stories/books.
- Tests—students take tests on comprehension, word meaning, and word pronunciation.

A direct instruction component of the program is designed to teach students how to identify main ideas and themes, draw conclusions, make predictions, and understand figurative language.

## Cost

The cost of implementing the *Talent Development Middle Grades Program* is available from the developer. The *Student Team Literature* component of the program costs approximately $400 per classroom. The curricular materials include teacher guides, partner discussion guides for students, literature tests, and word mastery tests for novels and plays. Teacher training begins with a two-day summer session, followed by monthly training throughout the year. Teachers in the program participate in monthly seminars to troubleshoot problems with instruction, extend their knowledge of the program's support materials, and enhance their skills in the program's instructional strategies.

## Research Summary

The WWC identified 18 studies that investigated the effects of the *Talent Development Middle Grades Program* on the reading achievement of adolescent readers.

The WWC reviewed all of those studies against group design evidence standards. No studies are randomized controlled trials that meet WWC evidence standards without reservations, and one study (Herlihy & Kemple, 2004) is a quasi-experimental design that meets WWC evidence standards with reservations. This study is summarized in this report. Three studies do not meet WWC evidence standards. The remaining 14 studies do not meet WWC eligibility screens for review in this topic area. Citations for all 18 studies are in the References section, which begins on p. 7.

**Table 2. Scope of reviewed research**

| | |
|---|---|
| **Grade** | 7, 8 |
| **Delivery method** | Whole class |
| **Program type** | Curriculum |
| **Studies reviewed** | 18 studies |
| **Group design studies that meet WWC evidence standards** | |
| • **without reservations** | 0 studies |
| • **with reservations** | 1 study |

### Summary of studies meeting WWC evidence standards without reservations

No studies of the *Talent Development Middle Grades Program* meet WWC evidence standards without reservations.

### Summary of study meeting WWC evidence standards with reservations

Herlihy and Kemple (2004) conducted a quasi-experiment that examined the effects of the *Talent Development Middle Grades Program* on students in grades 7 and 8 attending middle schools in an urban school district in the Northeast. Eleven intervention schools implemented the *Talent Development Middle Grades* comprehensive reform model. *Student Team Literature* was taught as the reading and English language arts curriculum. Eighteen comparison schools implemented the district's standard curriculum. Participating intervention and comparison schools were matched on school performance, ethnic composition, English proficiency, poverty, and student mobility, and each intervention school was matched with a set of comparison schools. This resulted in clusters of 1–12 comparison schools matched to each intervention school, with some comparison schools being matched to more than one intervention school. Of the 11 intervention schools, six used the program from 1997–98 to 2001–02, and five used the program in 2001–02. The authors refer to the first group as "early-implementing schools" and the latter group as "later-implementing schools"; we follow that naming convention in this report. An effect was measured after 1, 2, 3, 4, and 5 years of intervention implementation. Findings that reflect maximum exposure to the intervention by students were used to determine the rating of effectiveness in this WWC report.[6]

## Effectiveness Summary

The WWC review of the *Talent Development Middle Grades Program* for the Adolescent Literacy topic includes student outcomes in four domains: alphabetics, reading fluency, comprehension, and general literacy. The one study of the *Talent Development Middle Grades Program* that meets WWC evidence standards reported findings in one of the four domains: comprehension. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of the *Talent Development Middle Grades Program* on adolescent readers. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 16.

### Summary of effectiveness for the comprehension domain

One study reported findings in the comprehension domain.

Herlihy and Kemple (2004) found, and the WWC confirmed, a statistically significant positive effect of the *Talent Development Middle Grades Program* on the State Standards Assessment (SSA) in reading for eighth-grade students in year 2 in early-implementing treatment schools. Herlihy and Kemple (2004) did not find statistically significant positive effects of the *Talent Development Middle Grades Program* on the SSA in reading for eighth-grade students in years 1, 3, 4, and 5 or on the Stanford Achievement Test, 9th Edition, for year 5 seventh-grade students in early-implementing treatment schools. For later-implementing treatment schools in year 1, the authors did not find statistically significant positive effects of the *Talent Development Middle Grades Program* on the SSA in reading for eighth-grade students or on the Stanford Achievement Test, 9th Edition, for seventh-grade students (at $p < 0.05$). Because the effect on the SSA in reading for eighth-grade students in year 2 in early-implementing treatment schools is positive and statistically significant, and no effects are negative and statistically significant, the WWC characterizes these study findings as a statistically significant positive effect.

Thus, for the comprehension domain, one study showed statistically significant positive effects. This results in a domain rating of potentially positive effects, with a small extent of evidence.

### Table 3. Rating of effectiveness and extent of evidence for the comprehension domain

| Rating of effectiveness | Criteria met |
|---|---|
| **Potentially positive effects**<br>*Evidence of a positive effect with no overriding contrary evidence.* | In the one study that reported findings, the estimated impact of the intervention on outcomes in the *comprehension domain* was a statistically significant positive effect. |
| **Extent of evidence** | **Criteria met** |
| **Small** | One study that included an unspecified number of students in 29 schools reported evidence of effectiveness in the *comprehension domain*. |

## References

### Studies that meet WWC evidence standards without reservations

None

### Study that meets WWC evidence standards with reservations

Herlihy, C. M., & Kemple, J. J. (2004). *The Talent Development Middle School model: Context, components, and initial impacts on students' performance and attendance*. New York: MDRC.

***Additional source:***[7]

Herlihy, C. M., & Kemple, J. J. (2005). *The Talent Development Middle School model impacts through the 2002–2003 school year: An update to the December 2004 report.* New York: MDRC.

### Studies that do not meet WWC evidence standards

Mac Iver, D. J., Balfanz, R., Ruby, A., Byrnes, V., Lorentz, S., & Jones, L. (2004). Developing adolescent literacy in high poverty middle schools: The impact of Talent Development's reforms across multiple years and sites. In P. Pintrich & M. Maehr (Eds.), *Advances in motivation and achievement: Volume 13. Motivating students, improving schools: The legacy of Carol Midgley* (pp. 185–207). Oxford, England: Elsevier. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Mac Iver, D. J., Ruby, A., Balfanz, R., & Byrnes, V. (2002). Removed from the list: A comparative longitudinal case study of a reconstitution-eligible school. *Journal of Curriculum and Supervision, 18*(3), 259–289. The study does not meet WWC evidence standards because the measures of effectiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.

Plank, S. B., & Young, E. (2000). *Lessons for scaling up: Evaluations of the Talent Development Middle School's student team literature program. Report No. 46*. Baltimore, MD: Center for Research on the Education of Students Placed At Risk. The study does not meet WWC evidence standards because the measures of effectiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.

***Additional source:***

Mac Iver, D. J., Plank, S. B., & Balfanz, R. (1997). *Working together to become proficient readers: Early impact of the Talent Development Middle School's student team literature program. Report no. 15*. Baltimore, MD: Center for Research on the Education of Students Placed At Risk.

### Studies that are ineligible for review using the Adolescent Literacy Evidence Review Protocol

Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist, 42*(4), 223–235. The study is ineligible for review because it does not use a comparison group design or a single-case design.

***Additional source:***

Balfanz, R., Herzog, L., & Mac Iver, D. J. (n.d.) *Preventing student disengagement and keeping students on the graduation track in high-poverty middle-grades schools: Early identification and effective interventions.* Unpublished manuscript.

Balfanz, R., & Mac Iver, D. J. (2000). Transforming high-poverty urban middle schools into strong learning institutions: Lessons from the first five years of the Talent Development Middle School. *Journal of Education for Students Placed at Risk, 5*(1), 137–158. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2002). *Comprehensive school reform and student achievement: A meta-analysis*. Baltimore, MD: Center for Research on the Education of Students Placed at Risk. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Conklin, H. G. (2010). Preparing for the educational black hole? Teachers' learning in two pathways into middle school social studies teaching. *Theory & Research in Social Education, 38*(1), 48–79. The study is ineligible for review because it does not include a student outcome.

Juvonen, J., Le, V., Kaganoff, T., Augustine, C., & Constant, L. (2004). *Focus on the wonder years: Challenges facing the American middle school.* Santa Monica, CA: Rand Education. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Killion, J. (1999). *What works in the middle: Results-based staff development.* Oxford, OH: National Staff Development Council. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Mac Iver, D. J., Mac Iver, M. A., Balfanz, R., Plank, S. B., & Ruby, A. (2000). Talent Development Middle Schools: Blueprints and results for a comprehensive whole-school reform model. In M. G. Sanders (Ed.), *Schooling students placed at risk: Research, policy, and practice in the education of poor and minority adolescents* (pp. 261–288). Mahwah, NJ: Lawrence Erlbaum Associates Publishers. The study is ineligible for review because it a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Mac Iver, D. J., & Plank, S. B. (1996). *The Talent Development Middle School: Creating a motivational climate conducive to talent development in middle schools: Implementation and effects of Student Team Reading*. Baltimore, MD: Center for Research on the Education of Students Placed At Risk. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.

Mac Iver, D. J., Ruby, A., Balfanz, R., Jones, L., Sion, F., Garriott, M., & Byrnes, V. (2010). The Talent Development Middle Grades model: A design for improving early adolescents' developmental trajectories in high-poverty schools. In J. Meece & J. Eccles (Eds.), *Handbook of research on schools, schooling, and human development* (pp. 446–462). New York: Routledge. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Manzo, K. K. (2000). Finding their voices. *Education Week, 20*(5), 35–37. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Peterson, C. L., Caverly, D. C., Nicholson, S. A., O'Neal, S., & Cusenbary, S. (2001). *Building reading proficiency at the secondary level: A guide to resources.* Austin, TX: Southwest Educational Development Laboratory. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Quint, J. (2006). *Meeting five critical challenges of high school reform: Lessons from research on three reform models*. New York: MDRC. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Useem, E. (1998). *Teachers' appraisals of Talent Development Middle School training, materials, and student progress: Results from focus groups. Report no. 25.* Baltimore, MD: Center for Research on the Education of Students Placed At Risk. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Useem, E. L. (2001). *Second-year teachers' experience in Philadelphia's Talent Development Middle Schools*. Philadelphia, PA: Philadelphia Education Fund. The study is ineligible for review because it does not include a student outcome.

## Appendix A: Research details for Herlihy & Kemple, 2004

**Herlihy, C. M., & Kemple, J. J. (2004). *The Talent Development Middle School model: Context, components, and initial impacts on students' performance and attendance.* New York: MDRC.**

### Table A. Summary of findings

**Meets WWC evidence standards with reservations**

| Outcome domain | Sample size | Study findings | |
|---|---|---|---|
| | | Average improvement index *(percentile points)* | Statistically significant |
| **Comprehension** | 29 schools | +3 | Yes |

### Setting

The study took place in 29 middle schools in an urban school district in the Northeast.

### Study sample

Eleven *Talent Development (TD)* schools and 18 comparison schools participated in this quasi-experimental study. The *TD* schools used the *TD Student Team Literature* curriculum for reading and English language arts.[8] Each *TD* school was matched with a set of non-*TD* comparison schools that were similar on several dimensions, including racial/ethnic composition and math and reading test scores of eighth graders averaged over the 1995–96 and 1996–97 school years. This matching process resulted in groups (clusters) of 1–12 non-*TD* comparison schools for each *TD* school, with some non-*TD* schools serving as comparison schools for more than one *TD* school.

To estimate program impacts, two interrupted time series analyses were performed. The first compared the change in student test scores in *TD* schools *after* the program's implementation with the change of test scores of similar students in the *same* schools *prior* to *TD*'s implementation. The second interrupted time series analysis was conducted for the matched non-*TD* schools. Comparing the change in student test scores in *TD* schools to the change in student test scores in the matched non-*TD* schools produced program impact estimates. Specifically, the difference between deviations from the baseline in the *TD* schools and the deviations from the baseline in the non-*TD* schools on reading outcomes serves as the estimated program impact.

For the analysis of early-implementing *TD* schools (featured in Appendices C and D of this report), sample sizes were from two to six intervention schools and 18 comparison schools. For the analysis of later-implementing *TD* schools, sample sizes were five intervention schools and 18 comparison schools.

### Intervention group

The *Talent Development Middle Grades Program* includes a systematic reorganization of each school into small learning communities, organized around interdisciplinary teacher teams that share the same students and have common planning time. Teachers receive professional development on the use of the curriculum and accompanying instructional practice, and each school employs the services of curriculum coaches to help support teachers on an ongoing basis. The model provides catch-up opportunities during the school day for students who are struggling with mathematics or reading. Finally, the model facilitates school–family–community partnerships. The *TD* schools used *TD Student Team Literature* as their reading, English, and language arts curriculum, with most schools adopting it, at least partially, in the first year. The program includes partner discussion guides to assist students as they study fiction and non-fiction books and work in cooperative teams.

The early-implementing treatment schools used the *TD* model for five school years, from 1997–98 to 2001–02. The start of *TD* implementation was staggered across the six early-implementing schools, with some schools beginning *TD* implementation earlier than others. All six of the early-implementing schools had at least three years of implementation experience when the analysis was conducted: two schools had three years of experience, two schools had four years of experience, and two had implemented *TD* for five years. The later-implementing treatment schools used the *TD* model for one school year (2001–02, the most recent school year for which data are available in the analysis).

### Comparison group

Students in the comparison group received the district's standard reading/English language arts curriculum. The name of the standard district curriculum was not specified.

### Outcomes and measurement

The primary reading tests, administered annually in the school district, were the State Standards Assessment (SSA) and the Stanford Achievement Test, 9th Edition (SAT-9). The SSA was given in eighth grade, and the SAT-9 was given in seventh grade. Results for both tests were presented as Normal Curve Equivalent scores. For a more detailed description of these outcome measures, see Appendix B. The outcome measures that reflect the student's maximum exposure to the intervention are used to determine the WWC effectiveness rating and, therefore, are reported in Appendix C. The intermediate findings are reported in Appendix D.

### Support for implementation

The *Talent Development Middle Grades Program* provides four tiers of continuous support for teachers, including: (1) subject-specific professional development with a focus on modeling lessons, content knowledge, instructional strategies, and classroom management; (2) in-classroom support from a curriculum coach; (3) in-school support from teachers who receive extra training; and (4) support from instructional facilitators based at the Center for Research on the Education of Students Placed at Risk (CRESPAR).

## Appendix B: Outcome measures for the comprehension domain

| Comprehension | |
|---|---|
| *Stanford Achievement Test, 9th Edition (SAT-9)* | The SAT-9 is a standardized test used to measure academic knowledge of elementary and secondary school students in the United States. In this test, students answer multiple-choice questions on two reading subtests, reading vocabulary and reading comprehension. The scores from these two subtests were aggregated into a single total reading score. This test was given in seventh grade in the district, and results were presented as Normal Curve Equivalent scores in the study (as cited in Herlihy & Kemple, 2004). |
| *State Standards Assessment (SSA)–Reading* | This state standards assessment is a criterion-referenced test administered in grades 5, 8, and 11, which provides information about student performance on skills and content knowledge specified by the state. Results for this reading test were presented as Normal Curve Equivalent scores in the study (as cited in Herlihy & Kemple, 2004). |

## Appendix C: Findings included in the rating for the comprehension domain

| Outcome measure | Study sample | Sample size | Mean (standard deviation) Intervention group | Mean (standard deviation) Comparison group | WWC calculations Mean difference | WWC calculations Effect size | WWC calculations Improvement index | *p*-value |
|---|---|---|---|---|---|---|---|---|
| **Herlihy & Kemple, 2004[a]** | | | | | | | | |
| ***Early-implementing schools*** | | | | | | | | |
| *Three years of intervention* | | | | | | | | |
| *SSA–Reading* | Grade 8 Year 3 Cohort 3 | 6 clusters/ 24 schools | 29.10 (nr) | 29.00 (nr) | 0.10 | 0.01 | 0 | > 0.05 |
| *SSA–Reading* | Grade 8 Year 4 Cohort 4 | 4 clusters/ nr schools | 30.80 (nr) | 30.10 (nr) | 0.70 | 0.05 | +2 | > 0.05 |
| *SSA–Reading* | Grade 8 Year 5 Cohort 5 | 2 clusters/ nr schools | 35.50 (nr) | 32.70 (nr) | 2.80 | 0.17 | +7 | > 0.05 |
| *Two years of intervention* | | | | | | | | |
| *SSA–Reading* | Grade 8 Year 2 Cohort 2 | 6 clusters/ 24 schools | 30.50 (nr) | 27.80 (nr) | 2.70 | 0.17 | +7 | < 0.05 |
| *SAT-9* | Grade 7 Year 5 Cohort 6 | 2 clusters/ nr schools | 41.60 (nr) | 41.50 (nr) | 0.10 | 0.01 | 0 | > 0.05 |
| *One year of intervention* | | | | | | | | |
| *SSA–Reading* | Grade 8 Year 1 Cohort 1 | 6 clusters/ 24 schools | 27.70 (nr) | 28.80 (nr) | –1.10 | –0.07 | –3 | > 0.05 |
| ***Later-implementing schools*** | | | | | | | | |
| *One year of intervention* | | | | | | | | |
| *SSA–Reading* | Grade 8 Year 1 Cohort 7 | 5 clusters/ 23 schools | 32.40 (nr) | 31.70 (nr) | 2.70 | 0.05 | +2 | > 0.05 |
| *SAT-9* | Grade 7 Year 1 Cohort 8 | 5 clusters/ 23 schools | 40.70 (nr) | 37.70 (nr) | 3.00 | 0.17 | +7 | > 0.05 |
| **Domain average for comprehension (Herlihy & Kemple, 2004)** | | | | | | **0.07** | **+3** | **Statistically significant[b]** |

**Table Notes:** For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. Mean difference corresponds to the author-reported impact. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the average change expected for all students who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average student's percentile rank that can be expected if the student is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of the study's domain average was determined by the WWC. Years 1 to 5 are relative to the first calendar year of *Talent Development* implementation in each school (Herlihy & Kemple, 2004, p. 18). Cohorts 1–8 refer to non-overlapping student samples that received the *Talent Development Middle Grades Program*. nr=not reported. SSA = State Standards Assessment. SAT-9 = Stanford Achievement Test, 9th Edition.

[a] For Herlihy and Kemple (2004), no correction for clustering or multiple comparisons was needed. The *p*-values and effect sizes presented here were reported in the original study. The comparison group values are regression-adjusted estimates, which come from a multi-step process using ordinary least squares and controlling for fourth-grade math and reading SAT-9 test scores, race, and whether the student had repeated a prior grade. The intervention group values are comparison group values plus program impact calculated for a given year as the differences in deviation from the baseline average between intervention schools and comparison schools. For grade 8, the effect size was calculated by dividing the

impact in a given year by the standard deviation of that outcome for all eighth-grade students in the 18 comparison schools from preintervention school years 1995–96 and 1996–97. For grade 7, the effect size was calculated by dividing the impact in a given year by the standard deviation of that outcome for all seventh-grade students in the 18 comparison schools from preintervention school years 1995–96 and 1996–97.

[b] This study is characterized as having a statistically significant positive effect because the effect for at least one measure within the domain (SSA, grade 8, year 2) is positive and statistically significant, and no effects are negative and statistically significant. For more information, please refer to the *WWC Standards and Procedures Handbook*, version 2.1, p. 96.

## Appendix D: Summary of additional intermediate findings for the comprehension domain

| Outcome measure | Study sample | Sample size | Mean (standard deviation) | | WWC calculations | | | *p*-value |
|---|---|---|---|---|---|---|---|---|
| | | | Intervention group | Comparison group | Mean difference | Effect size | Improvement index | |
| **Herlihy & Kemple, 2004[a]** | | | | | | | | |
| ***Early-implementing schools*** | | | | | | | | |
| | | | *Two years of intervention* | | | | | |
| *SAT-9* | Grade 7 Year 2 Cohort 3 | 6 clusters/ 24 schools | 35.20 (nr) | 35.70 (nr) | –0.50 | –0.03 | –1 | > 0.05 |
| *SAT-9* | Grade 7 Year 3 Cohort 4 | 5 clusters/ 23 schools | 37.20 (nr) | 36.20 (nr) | 1.00 | 0.06 | +2 | > 0.05 |
| *SAT-9* | Grade 7 Year 4 Cohort 5 | 4 clusters/ nr schools | 37.50 (nr) | 35.90 (nr) | 1.60 | 0.09 | +4 | > 0.05 |
| | | | *One year of intervention* | | | | | |
| *SAT-9* | Grade 7 Year 1 Cohort 2 | 6 clusters/ 24 schools | 36.20 (nr) | 36.20 (nr) | 0.00 | 0.00 | 0 | > 0.05 |

**Table Notes:** The supplemental findings presented in this table are additional findings from the study in this report that do not factor into the determination of the effectiveness rating for the intervention. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. Mean difference corresponds to the author-reported impact. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the average change expected for all students who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average student's percentile rank that can be expected if the student is given the intervention. Years 1 to 4 are relative to the first calendar year of *Talent Development* implementation in each school (Herlihy & Kemple, 2004, p. 18). Cohorts 2–5 refer to non-overlapping student samples that received the *Talent Development Middle Grades Program*. nr = not reported. SAT-9 = Stanford Achievement Test, 9th Edition.

[a] For Herlihy and Kemple (2004), no correction for clustering or multiple comparisons was needed. The *p*-values and effect sizes presented here were reported in the original study. The comparison group values are regression-adjusted estimates, which come from a multi-step process using ordinary least squares and controlling for fourth-grade math and reading SAT-9 test scores, race, and whether the student had repeated a prior grade. The intervention group values are comparison group values plus program impact calculated for a given year as the differences in deviation from the baseline average between intervention schools and comparison schools. For grade 7, the effect size was calculated by dividing the impact in a given year by the standard deviation of that outcome for all seventh-grade students in the 18 comparison schools from preintervention school years 1995–96 and 1996–97.

## Endnotes

[1] The descriptive information for this program was obtained from publicly available sources: the program's website (http://www.csos.jhu.edu, downloaded July 2010) and the research literature (Mac Iver, Plank, & Balfanz, 1997). The WWC requests developers review the program description sections for accuracy from their perspective. The program description was provided to the developer in October 2011; however, the WWC received no response. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review. The literature search reflects documents publicly available by December 2011.

[2] The studies in this report were reviewed using the Evidence Standards from the WWC Procedures and Standards Handbook (version 2.1), along with those described in the Adolescent Literacy protocol (version 2.1). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

[3] The study that meets WWC evidence standards with reservations did not examine the effectiveness of the *Talent Development Middle Grades Program* on adolescent readers in the alphabetics, reading fluency, or general literacy achievement domains.

[4] For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 16. These improvement index numbers show the average and range of student-level improvement indices for all findings across the studies.

[5] The *Student Team Literature* program is an adaptation and elaboration of the *Student Team Reading and Writing* program (see WWC intervention report released in December 2011). Compared to *Student Team Reading and Writing, Student Team Literature* focuses on books rather than short selections from an anthology and includes a staff development component.

[6] The *Talent Development Middle Grades Program* was implemented in grades 6–8. The findings considered for the effectiveness rating reflect the maximum exposure of students to the program. For example, in the third, fourth, and fifth years of *Talent Development* implementation, eighth graders (from Cohorts 3, 4, and 5) had been exposed to the program over a period of three years (when they were in the sixth, seventh, and eighth grades). The corresponding intermediate findings (after two years of implementation) for seventh graders from the same Cohorts 3, 4, and 5 are reported in Appendix D and were not used for the rating of effectiveness. For Cohorts 2 and 6, two-year findings are considered for the effectiveness rating because these findings reflect the maximum exposure to the reading program for those students. For later-implementing schools (Cohorts 7 and 8), one-year findings (for seventh and eighth graders) are considered for the effectiveness rating because these findings reflect the maximum exposure to the reading program for those schools.

[7] Follow-up findings reported in Herlihy and Kemple (2005) are not included in this report because baseline equivalence was not demonstrated between intervention and comparison groups.

[8] One early-implementing school did not fully adopt *Student Team Literature* in all classrooms and eventually dropped it in the fifth year of *Talent Development* implementation. One of the later-implementing treatment schools never adopted the *Student Team Literature* program.

## Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, January). *Adolescent Literacy intervention report: Talent Development Middle Grades Program*. Retrieved from http://whatworks.ed.gov.

# WWC Rating Criteria

## Criteria used to determine the rating of a study

| Study rating | Criteria |
|---|---|
| **Meets WWC evidence standards without reservations** | A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT. |
| **Meets WWC evidence standards with reservations** | A study that provides weaker evidence for an intervention's effectiveness, such as a QED or an RCT with high attrition that has established equivalence of the analytic samples. |

## Criteria used to determine the rating of effectiveness for an intervention

| Rating of effectiveness | Criteria |
|---|---|
| **Positive effects** | Two or more studies show statistically significant positive effects, at least one of which met WWC evidence standards for a strong design, AND<br>No studies show statistically significant or substantively important negative effects. |
| **Potentially positive effects** | At least one study shows a statistically significant or substantively important positive effect, AND<br>No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects. |
| **Mixed effects** | At least one study shows a statistically significant or substantively important positive effect AND at least one study shows a statistically significant or substantively important negative effect, but no more such studies than the number showing a statistically significant or substantively important positive effect, OR<br>At least one study shows a statistically significant or substantively important effect AND more studies show an indeterminate effect than show a statistically significant or substantively important effect. |
| **Potentially negative effects** | One study shows a statistically significant or substantively important negative effect and no studies show a statistically significant or substantively important positive effect, OR<br>Two or more studies show statistically significant or substantively important negative effects, at least one study shows a statistically significant or substantively important positive effect, and more studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects. |
| **Negative effects** | Two or more studies show statistically significant negative effects, at least one of which met WWC evidence standards for a strong design, AND<br>No studies show statistically significant or substantively important positive effects. |
| **No discernible effects** | None of the studies shows a statistically significant or substantively important effect, either positive or negative. |

## Criteria used to determine the extent of evidence for an intervention

| Extent of evidence | Criteria |
|---|---|
| **Medium to large** | The domain includes more than one study, AND<br>The domain includes more than one school, AND<br>The domain findings are based on a total sample size of at least 350 students, OR, assuming 25 students in a class, a total of at least 14 classrooms across studies. |
| **Small** | The domain includes only one study, OR<br>The domain includes only one school, OR<br>The domain findings are based on a total sample size of fewer than 350 students, AND, assuming 25 students in a class, a total of fewer than 14 classrooms across studies. |

## Glossary of Terms

**Attrition** Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.

**Clustering adjustment** If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.

**Confounding factor** A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.

**Design** The design of a study is the method by which intervention and comparison groups were assigned.

**Domain** A domain is a group of closely related outcomes.

**Effect size** The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.

**Eligibility** A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

**Equivalence** A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.

**Extent of evidence** An indication of how much evidence supports the findings. The criteria for the extent of evidence levels are given in the WWC Rating Criteria on p. 16.

**Improvement index** Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from –50 to +50.

**Multiple comparison adjustment** When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.

**Quasi-experimental design (QED)** A quasi-experimental design (QED) is a research design in which subjects are assigned to intervention and comparison groups through a process that is not random.

**Randomized controlled trial (RCT)** A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into intervention and comparison groups.

**Rating of effectiveness** The WWC rates the effects of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 16.

**Single-case design** A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.

**Standard deviation** The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.

**Statistical significance** Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < 0.05$).

**Substantively important** A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the WWC Procedures and Standards Handbook (version 2.1) for additional details.