



WWC Single Study Review

A review of the design and summary of findings for an individual study



May 2016

WWC Review of the Report “Enhancing the Academic Development of Shy Children: A Test of the Efficacy of INSIGHTS”¹

The findings from this review do not reflect the full body of research evidence on the *INSIGHTS into Children’s Temperament* program.

What is this study about?

The study authors examined the effects of *INSIGHTS into Children’s Temperament (INSIGHTS)*, a program designed to enhance the development of students at risk for academic and behavioral difficulties in elementary school. During weekly meetings over a 10-week period, teachers, parents, and students are taught to recognize four temperament types, the strengths and weaknesses of each, and how to select strategies that are most appropriate for each child’s temperament.² *INSIGHTS* strives to improve the fit between students’ individual temperaments and the classroom environment, with the goal of enhancing behavioral engagement in classroom activities and ultimately academic outcomes.

Twenty-two elementary schools from three low-income, urban school districts were recruited to participate in the study and randomly assigned to the study conditions; 11 schools were assigned to receive the *INSIGHTS* program and the remaining 11 schools to the comparison group. Within those schools, study authors recruited 122 teachers from kindergarten and first-grade classrooms, and between 4 and 10 kindergarten students from each participating classroom, resulting in a total of 196 students in the intervention group and 178 students in the comparison group.

Baseline data collection occurred in the winter, after which schools were randomly assigned to receive either the *INSIGHTS* program or a supplementary afterschool reading program. Follow-up data on students’ academic and behavioral outcomes were collected at two

time points. The first follow up was conducted in the spring of kindergarten (1–3 months post-intervention). The sample for this analysis included 329 students (171 in the *INSIGHTS* group and 158 in the comparison group, Appendix C). The second follow up was conducted in the fall of first grade (6–8 months post-intervention).³ The sample for this analysis included 324 students (169 in the *INSIGHTS* group and 155 in the comparison group, Appendix D).⁴

The study authors examined the impact of *INSIGHTS* on student achievement using the Academic Competency Evaluation Scale (ACES). This measure assesses teachers’ perceptions of students’ skills relative to grade-level expectations in three areas: critical thinking, language arts, and mathematics. The study also examined impacts on students’ behavioral engagement in academic activities using the Behavioral Observation of Students in School (BOSS) tool.

WWC Rating

The research described in this report meets WWC group design standards without reservations

The study is a cluster randomized controlled trial with low levels of sample attrition at both the cluster (school) and subcluster (student) levels. This rating applies to the full sample of students in the 22 schools who received the *INSIGHTS* program or the supplementary afterschool reading program. This rating also applies to the subgroup of students classified as “shy” based on their score on the School-Aged Temperament Inventory.

What did the study find?⁵

In spring of kindergarten (1–3 months following the conclusion of the intervention), the WWC found that there were no statistically significant differences observed between students who received the *INSIGHTS* intervention and students in the comparison group on outcomes related to academic achievement and behavioral engagement. The WWC also found that, in fall of first grade, there were no statistically significant differences observed between students who received the *INSIGHTS* intervention and students in the comparison group on outcomes related to academic achievement and behavioral engagement. Similarly, in the subgroup of students who were classified as “shy” based on the School-Aged Temperament Inventory, the WWC found that there were no statistically significant differences observed on outcomes related to academic achievement in spring of kindergarten. However, in fall of first grade, the WWC found that there were statistically significant differences observed between “shy” students who received the *INSIGHTS* intervention and “shy” students in the comparison group on outcomes related to academic achievement.

Features of *INSIGHTS into Children's Temperament (INSIGHTS)*

INSIGHTS is a temperament-based intervention for teachers, parents, and children that targets early elementary school-age students who are at risk for academic and behavioral difficulties. The program aims to improve the fit between the classroom environment and students' individual temperaments to enhance their behavioral engagement and ultimately their academic outcomes. Parents and teachers are trained to match a student's temperament to one of four typologies, which can guide the choice of strategies for learning and self-regulation, and to use a “scaffold-and-stretch approach” when students encounter challenges. The program lasts 10 weeks and has the following components:

- (1) Training sessions for parents and teachers: During weekly sessions, caregivers learn to recognize a student's temperament typology and use strategies that match the student's temperament. For this study, one training session was jointly held with teachers and parents, and the others were delivered separately, as teacher sessions included classroom management and parent sessions included parenting skills.
- (2) Classroom sessions for students: Led by a program facilitator, study sessions lasted 45 minutes and focused on empathy and self-regulation using puppets, workbooks, flash cards, and videos.

Appendix A: Study details

O'Connor, E. E., Cappella, E., McCormick, M. P., & McClowry, S. G. (2014). Enhancing the academic development of shy children: A test of the efficacy of INSIGHTS. *School Psychology Review, 43*(3), 239–259.

Setting The study was conducted in kindergarten classrooms in 22 elementary schools in three low-income, urban school districts.

Study sample Twenty-two elementary schools were recruited to participate in the randomized controlled trial. This study reports findings for the first two follow-up periods. Schools were randomly assigned to receive either the *INSIGHTS* program or a supplementary afterschool reading program. The study authors recruited 122 teachers from kindergarten and first-grade classrooms in participating schools, and between four and 10 kindergarten students from each participating classroom. Baseline data collection occurred in the winter, after which schools were randomly assigned to receive either the *INSIGHTS* program or a supplementary afterschool reading program. At the time of randomization, there were 374 kindergarten students in the study; 196 students in 11 schools were assigned to receive the *INSIGHTS* program, and 178 in the remaining 11 schools were assigned to the comparison group. Follow-up data on students' academic and behavioral outcomes were collected at two time points: (1) spring of kindergarten (May/June; 1–3 months post-intervention), and (2) fall of first grade (October/November; 6–8 months post-intervention). The analytic sample in spring of kindergarten included 329 students (171 in the *INSIGHTS* group and 158 in the comparison group). There were 324 students in the analytic sample in the fall of first grade (169 in the *INSIGHTS* group and 155 in the comparison group). Of the 345 students in the analytic sample with valid data on at least one follow-up assessment (spring of kindergarten or fall of first grade), 91% were Black or Hispanic, and 87% qualified for free or reduced-price lunch programs. This WWC review focuses on outcomes measured in the spring of kindergarten, the time period closest to the end of the intervention. In response to an author query, the WWC confirmed that the study reports estimates for the sample of students with complete outcome data from the spring of kindergarten.

Intervention group *INSIGHTS* is a temperament-based program to support the development of low-income students at risk for academic and behavioral difficulties in elementary school. The program aims to improve the fit between the classroom environment and students' individual temperaments (as measured by the School-Aged Temperament Inventory) to enhance their behavioral engagement and ultimately their academic outcomes. Parents and teachers are trained to match a student's temperament to one of four typologies, which can guide the choice of strategies for learning and self-regulation, and to use a "scaffold-and-stretch approach" when students encounter challenges. During weekly meetings over a 10-week period, teachers, parents, and students are taught to recognize four temperament types, the strengths and weaknesses of each, and how to select strategies to match temperaments. Student sessions were 45 minutes long, were conducted in the classroom, and focused on empathy and self-regulation using puppets, workbooks, flash cards, and videos. Teachers and parents attended 2-hour meetings focused on increasing responsiveness to various temperaments using a structured curriculum. One meeting was jointly held with parents and teachers, and the others were separate teacher sessions focused on classroom management and parent sessions focused on parenting skills. Both teachers and parents received incentives for attending.

Comparison group

Students in the comparison schools participated in a 10-week afterschool reading program, and their teachers attended two early literacy workshops, which were also presented to parents. As with the intervention group, both teachers and parents received incentives for attending.

Outcomes and measurement

Study authors examined three outcomes from the Academic Competency Evaluation Scale (ACES) (DiPerna & Elliott, 2000). The outcomes are subscales of teacher perceptions of the students' skills relative to grade-level expectations in their school in three areas: critical thinking, language arts, and mathematics. The study also assessed students' behavioral engagement in academic activities using the Behavioral Observation of Students in School (BOSS) tool. Both the ACES and BOSS were administered at three points in time: baseline (January/February of kindergarten), first follow up (May/June kindergarten; 1–3 months post-intervention), and second follow up (October/November of first grade; 6–8 months post-intervention). Because the first follow-up outcomes were measured closest to the end of the intervention, they were the primary outcomes for this study. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

The eight *INSIGHTS* facilitators attended a semester-long graduate course before conducting the intervention. Each school was assigned one facilitator who worked with the teachers, parents, and students at that school. Facilitators followed scripts, used checklists, and documented sessions to maintain fidelity to the program model. Facilitators participated in weekly supervision sessions with the program developer, in which they discussed challenges, implementation logistics, participant concerns, and deviations from the program model.

Reason for review

This study was identified for review by the WWC because it was supported by a grant (R305A080512) to New York University (Principal Investigator: Sandee McClowry) from the National Center for Education Research (NCER) at the Institute of Education Sciences (IES).

Appendix B: Outcome measures for each domain

Academic achievement	
<i>Academic Competency Evaluation Scale (ACES) Critical Thinking subscale</i>	The ACES Critical Thinking subscale measures teacher perceptions of students' critical thinking skills relative to grade-level expectations (DiPerna & Elliott, 2000). Critical thinking skills are rated on a 5-point scale ranging from "far below" to "far above" expectations. The Critical Thinking subscale includes nine items asking teachers to report how well students engage in reflection, analysis, synthesis, and investigation. The average internal consistency for this subscale across all three time points was .97.
<i>ACES Language Arts subscale</i>	The ACES Language Arts subscale measures teacher perceptions of students' language arts skills relative to grade-level expectations (DiPerna & Elliott, 2000). Language arts skills are rated on a 5-point scale ranging from "far below" to "far above" expectations. The Language Arts subscale includes 11 items about skills necessary for generating and understanding written language, including reading comprehension and written communication. The average internal consistency for this subscale across all three time points was .97.
<i>ACES Mathematics subscale</i>	The ACES Mathematics subscale measures teacher perceptions of students' mathematics skills relative to grade-level expectations (DiPerna & Elliott, 2000). Mathematics skills are rated on a 5-point scale ranging from "far below" to "far above" expectations. The Mathematics subscale includes eight items about skills related to the use of numbers, including measurement, computation, and problem solving. The average internal consistency for this subscale across all three time points was .98.
Behavioral engagement	
<i>Behavioral Observation of Students in Schools (BOSS)</i>	The BOSS is an observational tool used to assess students' behavioral engagement in academic activities. Trained staff conducted classroom observations of students and record the amount of time that students are actively engaged in academic activities. Observations were conducted during two 15-minute intervals on separate days. Inter-rater reliability ranged from .80 to .95.

Table Notes: Additional data on the ACES critical thinking outcome measures were collected from the following source: DiPerna, J. C., Volpe, R. J., & Elliott, S. N. (2005). A model of academic enablers and mathematics achievement in the elementary grades. *Journal of School Psychology, 43*(5), 379–392.

Appendix C: Study findings for each domain—estimates from spring of kindergarten (1–3 months post-intervention)

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Academic achievement								
<i>Academic Competency Evaluation Scale (ACES) Critical Thinking subscale</i>	Full sample, spring of kindergarten	22 schools/ 329 students	2.82 (0.62)	2.95 (0.58)	-0.13	-0.22	-9	nr
<i>ACES Language Arts subscale</i>	Full sample, spring of kindergarten	22 schools/ 329 students	2.80 (0.72)	2.89 (0.69)	-0.09	-0.13	-5	nr
<i>ACES Mathematics subscale</i>	Full sample, spring of kindergarten	22 schools/ 329 students	2.79 (0.65)	2.91 (0.58)	-0.12	-0.19	-8	nr
Domain average for academic achievement						-0.18	-7	Not statistically significant
Behavioral engagement								
<i>Behavioral Observation of Students in Schools (BOSS)</i>	Full sample, spring of kindergarten	22 schools/ 324 students	0.69 (0.26)	0.66 (0.22)	0.03	0.12	+5	nr
Domain average for behavioral engagement						0.12	+5	Not statistically significant

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. The unadjusted means and standard deviations displayed in this table were obtained through a response to an author query. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals 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 individual’s percentile rank that can be expected if the individual 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. Some statistics may not sum as expected due to rounding. nr = not reported.

Study Notes: The study authors present results of statistical tests based on analyses that are not accepted by the WWC; therefore, the author-reported p-values are not displayed in the table. The WWC-computed p-values for the statistical tests comparing the three unadjusted posttest measures of academic achievement are as follows: .49 for the ACES Critical Thinking measure, .29 for the ACES Language Arts measure, and .44 for the ACES Mathematics measure. A correction for multiple comparisons was needed for the three measures of academic achievement but did not affect whether any of the contrasts were found to be statistically significant. This study is characterized as having an indeterminate effect in the academic achievement domain because the mean effect reported is neither statistically significant nor substantively important. The WWC-computed p-value for the BOSS is .27. This study is characterized as having an indeterminate effect in the behavioral engagement domain because the mean effect reported is neither statistically significant nor substantively important. For more information, please refer to the WWC Standards and Procedures Handbook (version 3.0), p. 27.

Appendix D.1: Supplemental findings by domain—estimates for full sample from fall of first grade (6–8 months post-intervention)

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Academic achievement								
<i>Academic Competency Evaluation Scale (ACES) Critical Thinking subscale</i>	Full sample, fall of first grade	22 schools/ 324 students	2.56 (0.65)	2.58 (0.65)	-0.02	-0.03	-1	nr
<i>ACES Language Arts subscale</i>	Full sample, fall of first grade	22 schools/ 324 students	2.49 (0.89)	2.46 (0.80)	0.03	0.04	+1	nr
<i>ACES Mathematics subscale</i>	Full sample, fall of first grade	22 schools/ 324 students	2.52 (0.73)	2.63 (0.67)	-0.11	-0.16	-6	nr
Behavioral engagement								
<i>Behavioral Observation of Students in Schools (BOSS)</i>	Full sample, fall of first grade	22 schools/ 315 students	0.69 (0.15)	0.73 (0.15)	-0.04	-0.27	-10	nr

Table Notes: The supplemental findings presented in this table are additional findings that meet WWC design standards without reservations, but do not factor into the determination of the study rating. The unadjusted means and standard deviations displayed in this table were obtained through a response to an author query. 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. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals 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 individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding. nr = not reported.

Study Notes: For the overall sample, the study authors present results of statistical tests based on analyses that are not accepted by the WWC; therefore, the author-reported p-values are not displayed in the table. The WWC-computed p-values for the statistical tests comparing the three measures of academic achievement are as follows: .78 for the ACES Critical Thinking measure, .75 for the ACES Language Arts measure, and .16 for the ACES Mathematics measure. A correction for multiple comparisons was needed for the three measures of academic achievement but did not affect whether any of the contrasts were found to be statistically significant. The WWC-computed p-value for the BOSS is .02. For more information, please refer to the WWC Standards and Procedures Handbook (version 3.0), p. 27.

Appendix D.2: Supplemental findings for “shy” subgroup in spring of kindergarten (1–3 months post-intervention)

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Academic achievement								
<i>ACES Critical Thinking subscale</i>	“Shy” subsample, spring of kindergarten	22 schools/ 101 students	2.38 (0.64)	2.26 (0.61)	0.12	0.19	+8	nr
<i>ACES Language Arts subscale</i>	“Shy” subsample, spring of kindergarten	22 schools/ 101 students	2.50 (0.75)	2.53 (0.71)	−0.03	−0.04	−2	nr
<i>ACES Mathematics subscale</i>	“Shy” subsample, spring of kindergarten	22 schools/ 101 students	2.66 (0.67)	2.61 (0.62)	0.05	0.08	+3	nr

Table Notes: The supplemental findings presented in this table are additional findings that meet WWC design standards without reservations, but do not factor into the determination of the study rating. The means displayed in this table reflect adjusted means obtained through a response to an author query. 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. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals 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 individual’s percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding. nr = not reported.

Study Notes: For the “shy” subgroup sample, the study authors present results of statistical tests based on analyses that are not accepted by the WWC; therefore, the author-reported p-values are not displayed in the table. The WWC-computed p-values for the statistical tests comparing the three measures of academic achievement in spring of kindergarten are as follows: .38 for the ACES Critical Thinking measure, .84 for the ACES Language Arts measure, and .82 for the ACES Mathematics measure. A correction for multiple comparisons was needed for the three measures of academic achievement but did not affect whether any of the contrasts were found to be statistically significant. For more information, please refer to the WWC Standards and Procedures Handbook (version 3.0), p. 27.

Appendix D.3: Supplemental findings for “shy” subgroup in fall of first grade (6–8 months post-intervention)

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Academic achievement								
<i>ACES Critical Thinking subscale</i>	“Shy” subsample, fall of first grade	22 schools/99 students	2.48 (0.59)	2.15 (0.62)	0.33	0.54	+21	nr
<i>ACES Language Arts subscale</i>	“Shy” subsample, fall of first grade	22 schools/99 students	2.49 (0.81)	2.51 (0.79)	-0.02	-0.02	-1	nr
<i>ACES Mathematics subscale</i>	“Shy” subsample, fall of first grade	22 schools/99 students	2.67 (0.67)	2.44 (0.69)	0.23	0.34	+13	nr

Table Notes: The supplemental findings presented in this table are additional findings that meet WWC design standards without reservations, but do not factor into the determination of the study rating. 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. The means displayed in this table reflect adjusted means obtained through a response to an author query. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals 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 individual’s percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding. nr = not reported.

Study Notes: For the “shy” subgroup sample, the study authors present results of statistical tests based on analyses that are not accepted by the WWC; therefore, the author-reported p-values are not displayed in the table. The WWC-computed p-values for the statistical tests comparing the three measures of academic achievement in fall of first grade are as follows: .01 for the ACES Critical Thinking measure, .90 for the ACES Language Arts measure, and .10 for the ACES Mathematics measure. A correction for multiple comparisons was needed and resulted in a WWC-computed critical p-value of .02 for the ACES Critical Thinking measure; therefore, the WWC finds the result to be statistically significant. For more information, please refer to the WWC Standards and Procedures Handbook (version 3.0), p. 26.

Endnotes

¹ Single study reviews examine evidence published in a study (supplemented, if necessary, by information obtained directly from the authors) to assess whether the study design meets WWC group design standards without reservations. The review reports the WWC's assessment of whether the study meets WWC group design standards without reservations and summarizes the study findings following WWC conventions for reporting evidence on effectiveness. This study was reviewed using the single study review protocol (version 2.0).

² The program uses characters associated with the four empirically derived temperament typologies, which are: (1) Hilary the Hard Worker (industrious); (2) Gregory the Grumpy (high maintenance); (3) Fredrico the Friendly (social and eager to try); and (4) Coretta the Cautious (shy).

³ The article does not specify the amount of time post-intervention at which the spring kindergarten and fall first-grade assessments occurred. The WWC computed this amount of time based on the information provided in the study about the dates of the three assessments (January/February, May/June, and October/November) and the length of the intervention (10 weeks).

⁴ This review focuses on outcomes measured in the full sample of students during spring of kindergarten, the time period closest to the end of the intervention.

⁵ In the study report, the authors present results of statistical tests based on analyses that use an imputation method that is not accepted by the WWC. The analysis also did not present separate results for spring of kindergarten and fall of first grade; rather, the authors reported on the results of a longitudinal analyses containing an interaction term of treatment by time. In a response to an author query, the authors provided unadjusted means and standard deviation for the sample of students with unimputed data at all three time points. Statistical tests were performed by the WWC.

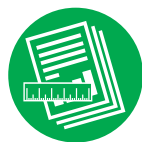
Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2016, May). *WWC review of the report: Enhancing the academic development of shy children: A test of the efficacy of INSIGHTS*. Retrieved from <http://whatworks.ed.gov>

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 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 analytic sample groups are similar on observed characteristics defined in the review area protocol.
Improvement index	Along a percentile distribution of individuals, the improvement index represents the gain or loss of the average individual due to the intervention. As the average individual 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 study participants 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 eligible study participants are randomly assigned to intervention and comparison groups.
Single-case design (SCD)	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 are 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% ($p < .05$).
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 \(version 3.0\)](#) for additional details.



Intervention
Report



Practice
Guide



Quick
Review



Single Study
Review

A **single study review** of an individual study includes the WWC's assessment of the quality of the research design and technical details about the study's design and findings.

This single study review was prepared for the WWC by Mathematica Policy Research under contract ED-IES-13-C-0010.