Core Academic Language Skills Instrument (CALS-I): Final Report

Paola Uccelli, *Harvard Graduate School of Education*
Christopher D. Barr, *University of Houston*
Emily Phillips Galloway, *Harvard Graduate School of Education*
August, 2016

Acknowledgment
The research reported here was supported by the Institute of Education Sciences of the U.S. Department of Education through grant R305F100026, which was awarded to the Strategic Education Research Partnership as part of the Reading for Understanding Research Initiative. The opinions expressed are those of the authors and do not represent the views of the institute or the U.S. Department of Education. We express our gratitude to the students and teachers who shared their valuable time and insights with us and to our numerous colleagues for their helpful comments as we conducted this work.

CALS-I | Purpose

The *Core Academic Language Skills Instrument* (CALS-I) was designed to measure high-utility academic language skills hypothesized to support reading comprehension across the content areas in grades 4 through 8. This assessment is guided by the hypothesis that the variability in the identified set of CALS is relevant to predict reading comprehension for bilingual and monolingual students in the upper elementary and middle school years. As a complement to measures of academic vocabulary knowledge and discipline-specific skills, the main purpose of the CALS-I is to measure *a comprehensive set of high-utility language* skills that are *relevant across content areas*. (Uccelli, Barr, Dobbs, Phillips Galloway, Meneses, & Sánchez, 2015; Uccelli, Phillips Galloway, Barr, Meneses, & Dobbs, 2015).

CALS-I | Description

The CALS-I is a group-administered instrument designed to assess core academic language skills (CALS) in grades 4 to 8. Each CALS-I form consists of a 50-minute paper-and-pencil test that includes eight tasks: *Connecting Ideas*, *Tracking Themes*, *Organizing Texts*, *Breaking Words*, *Comprehending Sentences*, *Identifying Definitions*, *Interpreting Epistemic Stance Markers*, and *Understanding Metalinguistic Vocabulary.* A ninth optional task assesses degree of adherence to academic register expectations through short open-ended responses. Tasks assess students' skills through a range of multiple choice, matching, or short written responses. The CALS-I has two forms:

- CALS-I-Form 1 for 4^{th}-6^{th} grade (α =.90)
- CALS-I-Form 2 for 7^{th}-8^{th} grade (α =.86)

The CALS-I is vertically equated and normed for English proficient students attending urban public schools in grades 4-8.

CALS | Operational definition

Core Academic Language Skills (CALS) refer to a constellation of skills that correspond to linguistic features prevalent in academic texts, yet rare in colloquial conversations (e.g., knowledge of logical markers that connect ideas, such as *nevertheless, consequently;* knowledge of structures that pack dense information, such as nominalizations or embedded clauses; knowledge of structures for organizing analytic texts). This constellation of skills was hypothesized to support academic reading across school content areas. As shown in Figure 1, the CALS construct includes seven domains (see Appendix 1 for CALS-I Task Descriptions).

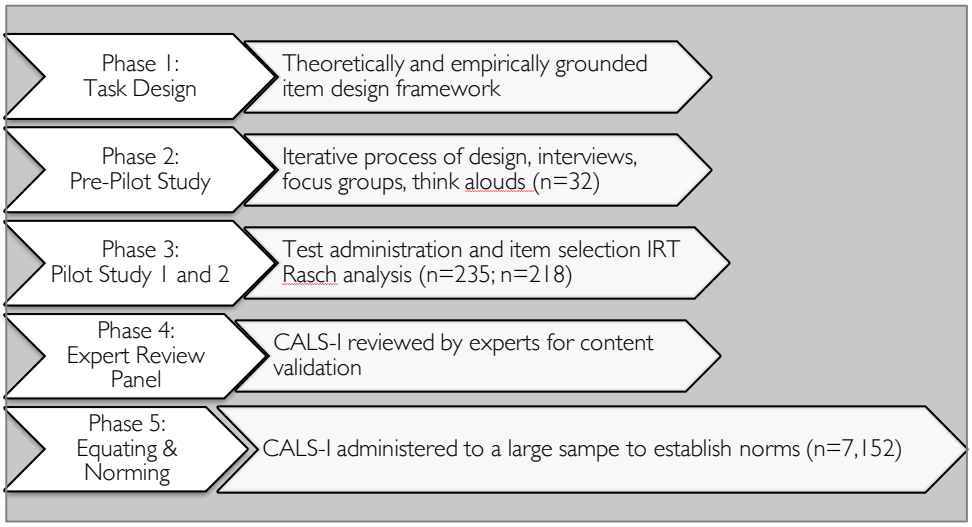
Figure 1. CALS-I: Domains and skills measured

CALS Domain	Skills measured
Unpacking dense information	Skill in comprehending and using complex words and complex sentences that facilitate concise communication (e.g., nominalizations, embedded clauses, expanded noun phrases).
Connecting ideas logically	Skill in comprehending and using 'connectives' prevalent in academic texts to signal relationships between ideas (e.g., *consequently, on the one hand…on the other hand).*
Tracking participants and ideas	Skill in identifying or producing the terms or phrases used to refer to the same participants or themes throughout an academic text (e.g., *Water evaporates at 100 degrees Celsius. This process…).*
Organizing analytic texts	Skill in organizing analytic texts, especially argumentative texts, according to its conventional academic structure (e.g., thesis, argument, counterargument, conclusion) and paragraph-level structures (e.g., compare/contrast; problem/solution)
Understanding metalinguistic vocabulary	Skill in understanding precise meanings, in particular, in using language to make thinking and reasoning visible, known as metalinguistic vocabulary (e.g., *hypothesis, generalization, argument)*
Understanding a writer's viewpoint	Skill in understanding or using markers that signal a writer's viewpoint, especially a 'epistemic stance markers', those that signal a writer's degree of certainty in relationship to a claim (e.g., *certainly; it is unlikely that)*
Recognizing academic language	Skill in recognizing more academic language when contrasted with more colloquial language in communicative contexts where academic language use is expected (e.g., more *academic vs. more colloquial definitions of nouns)*

CALS-I | Development & Validation

The development of this instrument included expert linguists, psychologists, psychometricians, and educators and was guided by theoretical and empirical research. The iterative design process unfolded in the following sequence: a task design phase, and pre-pilot study, a series of qualitative and quantitative pilot studies, an expert review panel, and a norming phase.

Figure 3. CALS-I Development Phases.



Setting and Participants: All 4th-8th grade participants in the validation studies so far conducted attended large urban middle schools in the Northeast and South East regions of the United States. School samples were representative of their respective school districts populations (see Appendix 2, Tables 1 ,2, 3 display socio-demographic characteristics of for samples that informed the CALS-I development).

Data Collection and Analysis: Data were collected in a series of studies by trained research assistants. Classical Test Theory (CTT) and Rasch Item Response Theory (IRT) analyses were used to inform the design of two forms of the instrument. To determine if items were a good fit, the point biseral correlations and, as well as infit and outfit statistics were examined. Additionally, the relation between item level performance and the criterion were examined. To investigate the dimensionality of the CALS-I factor we started by examining if the first residual Rasch factor added any incremental criterion validity to the Rasch factor and then followed this analysis with confirmatory factor analytic methods.

Research Design: The design and validation of the CALS-I followed 5 phases:

1. **Task Design Phase | Task and Item Design Specifications.** Guided by theory and developmental research, the first phase involved designing items that would

capture academic language skills while minimizing the impact of decoding, vocabulary knowledge, prior knowledge, productive writing demands, and extended text comprehension demands. To overcome the logistical difficulties of individual language assessment (a procedure that schools find disruptive), we designed items that could be group-administered. Item design was guided by a developmental map built on the basis of extensive literature review on textual linguistics, developmental linguistics, and educational linguistics research.

2. **Pre-Pilot Phase | Iterative Item Testing and Refinement**. The design of items followed an iterative process of generation, testing, incorporation of students' feedback obtained through multi-party interviews using a structured protocol, and retesting that resulted in modifying, recalibrating, or discarding individual items and in improving task instructions. All items were piloted in individual or small-group interviews with a sample of 32 students (grades 4-8).

3. **Pilot Phase | Selection of CALS-I Final Item set**. The third phase consisted of two pilot studies that informed the selection of items that comprise the CALS-I. Using CTT and Rasch IRT modeling in **Study 1**, an initial set of items was selected from a larger pool of items (a total of 130 items), administered to a cross-sectional sample of 4^{th} to 8^{th} graders (n=235) (Table 1 displays the Study 1 sample characteristics). Reliability was investigated and found to be robust at .92 as indexed by coefficient alpha and at .82 by split-half reliability of even vs. odd numbered items. Additionally, when comparing the Test Information Function to the Standard Error of Measurement, the results indicated that the Test Information Function was more than two times greater than the Standard Error of Measurement, even at +/- two SD. A confirmatory factor analysis (CFA) applied to the final set of items suggested a single factor: (CFI=.93, TLI = .92, RMSEA <.05). The zero order within-grade correlations between the CALS-I total score and the Massachusetts Comprehensive Assessment (MCAS-ELA) (a non-vertically-equated state-wide assessment) were statistically significant and ranged from .41 for grade 7 to .77 for grade 6. Multiple linear regression analyses within-grade revealed that students' CALS-I scores were significant predictors of reading comprehension scores (as measured by the statewide standardized ELA assessment), even after controlling for reading fluency (Uccelli, Barr, et al., 2015). In pilot **Study 2** (n= 218 4^{th}-6^{th} graders; 139 7^{th} & 8^{th} graders), two additional tasks were incorporated to the CALS-I: (a) knowledge of epistemic stance markers (e.g., *it is possible, likely)* and (b) knowledge of academic metalinguistic vocabulary, i.e. words that denote language or thinking processes and support text-based discussion and argumentation *(counterargument, generalization)*. In addition, a second pilot form was generated for grades 7 and 8 to address ceiling effects detected in Study 1 for these older students (Table 2 displays the Study 2 sample characteristics for 4^{th}-6^{th} graders). Confirming prior results, a confirmatory factor model fitting CALS-I items to a single factor produced good model fit offering evidence of unidimensionality (RMSEA = .06, CFI = .95, and TLI =

.94)[1]. For Pilot Form 1, reliability was .93 as indexed by coefficient alpha and .90 by split half reliability. Validity was .70 as indexed by the zero order correlation with the Gates-MacGinitie Passage Comprehension for the total sample. For Pilot Form 2, reliability was .91 as indexed by coefficient alpha and .84 by split half reliability. Validity was .75 as indexed by the zero order correlation with the Gates-MacGinitie Passage Comprehension for the total sample. Data from this cross-sectional sample revealed that, beyond the contribution of reading fluency and academic vocabulary knowledge, CALS-I scores significantly contributed to predict reading comprehension as measured by the Gates-McGinitie reading comprehension test (Uccelli, Phillips Galloway, et al., 2015). Thus, preliminary findings showed promising results for the reliability and validity of this extended form. However, some items were modified as the result of these analyses, so the evidence of reliability and validity information for the finalized forms was produced in Phase 5 (Validation, Equating & Norming) and on the basis of data from a larger sample.

4. **Expert Panel Review Phase | CALS-I Content Validation.** The fourth phase, feedback from experts in the field was solicited to establish content validity. The CALS-I along with a content validation survey was sent to five experts in the field of academic language for their independent review. Overall, the team of experts gave the CALS-I a mean score of 3.5 out of 5-point scale, with a median score of 4. Experts' valuable feedback and recommendations were, to the extent possible, incorporated in the assessment.

5. **Validation, Equating & Norming Phase | Vertically equated CALS-I**. In this fifth phase, again guided by a combination of CTT and IRT analyses, as well as theoretical rationales for item inclusion, the two CALS-I forms were finalized (the final forms remained fairly consistent with those of Study 2 because only two items were deleted). From a cross-sectional and longitudinal sample of CALS-I data collected across three years in our project, for the validation, scaling and norming we selected a subset of the data that included only CALS-I scores for students who were administered the CALS-I *for the first time* throughout the three years of our project. We focused only on the first instance of the CALS-I administration to prevent any potential retesting effect from impacting the psychometric results. This so-called *calibration data set* consisted of a cross-sectional sample of a total of 7,152 students from grades 4 to 8. Following the creation of the calibration data set, we ran our scoring Rasch model with anchored item difficulties and extracted factor scores. Reliability and validity evidence for the final set of items for Form 1 and Form 2 was investigated and found to be robust. For Form 1, reliability was .90 as indexed by coefficient alpha and validity was .69 as indexed by the zero order correlation with the ETS-developed reading comprehension assessment *Global,*

[1] Additionally, a model relating individual factor score for the CALS-I factor and the first residual factor to Gates-MacGinitie indicated that the residual factor did not provide significant incremental prediction of the Gates-MacGinitie – further evidence of unidimensionality.

Integrated, Scenario-based Assessment (GISA). For Form 2, reliability was .86 as indexed by coefficient alpha and validity was .71 as indexed by the zero order correlation with the GISA. , CFA factor models were examined and found that single factor (RMSEA = .02, CFI = .97, and TLI = .97)[2].

For norming a series of steps were followed on the basis of the calibration data set analysis. Given that the full range of factor scores includes negative values, these negative scores could be interpreted erroneously as poor performance by test administrators and test takers. Thus, to avoid such misinterpretation, factor scores were rescaled to a positive metric and Extended CALS scores (ECALS scores) were generated with a mean ECALS of 500 and a standard deviation of 50. Form-specific tables that display the full range of possible raw CALS-I scores and their corresponding CALS-I factor scores and ECALS scores were produced. These tables could be used by examiners after administering the assessment, such that once students' total raw scores are calculated, an ECALS score can be assigned for each student based on these tables.

CALS-I | Psychometric Properties

After psychometric analysis and theoretical considerations, two final assessment forms were created, Form 1 for grades 4-6 and Form 2 for grades 7-8. Form 1 consists of 49 items and Form 2 consists of 46 items, with 30 overlapping items across forms. Reliability and validity evidence were robust. For Form 1, reliability was .90 as indexed by coefficient alpha. Validity was .69 as indexed by the zero order correlation with the GISA for the total Form 1 sample. For Form 2, reliability was .86 as indexed by coefficient alpha. Validity was .71 as indexed by the zero order correlation with the GISA for the total Form 2 sample. This is consistent with the robust reliability and validity evidence of Study 2 in which the Gates MacGinitie was used as criterion. CFA factor models were examined and found that a single factor model fit the data well. In sum, the final two forms of the now vertically equated CALS-I have strong evidence of reliability, as well as construct and criterion validity, and now include norming information to guide the interpretation of results using the ECALS scores

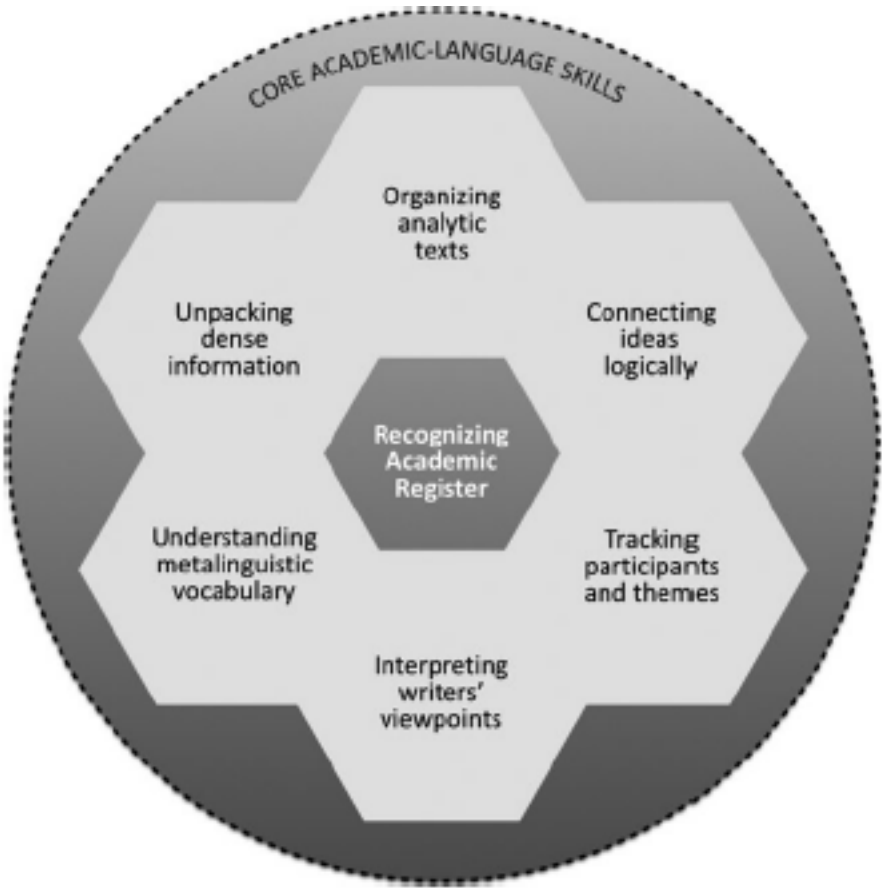
CALS-I | Potential Uses

The CALS-I is presently available for use as a research instrument upon request. Results of the CALS-I have been used effectively in teachers' professional development to raise awareness of the importance of paying attention to core academic language skills during instruction. Additional uses of the CALS-I to inform pedagogical practice are

[2] Additionally, a model relating individual factor score for the CALS-I factor and the first residual factor to Gates-MacGinitie indicated that the residual factor did not provide significant incremental prediction of the Gates-MacGinitie – further evidence of unidimensionality.

currently being investigated with encouraging preliminary results (Uccelli, Phillips Galloway, Aguilar, & Allen, 2016). Currently, norms for the CALS-I are available for English proficient students enrolled in U.S. public urban schools in grades 4 to 8.

Figure 3. Visual representation of CALS construct.



Note. From "Beyond Vocabulary: Exploring Cross-Disciplinary Academic-Language Proficiency and Its Association With Reading Comprehension," by P. Uccelli, E. Phillips Galloway, C.D. Barr, A. Meneses, and C.L. Dobbs, 2015, *Reading Research Quarterly*, *50*(3), p. 349. Copyright 2015 by the International Literacy Association. Reprinted with permission.

References

Uccelli, P., Phillips Galloway, E., Barr, C., Meneses, A., & Dobbs, C. (2015). *Beyond vocabulary: Core Academic Language Skills (CALS) that support text comprehension. Reading Research Quarterly, 50*(3), 261-359. doi: 10.1002/rrq.104

Uccelli, P., Barr, C. D., Dobbs, C. L., Phillips Galloway, E., Meneses, A., & Sánchez, E. (2015). Core academic language skills (CALS): An expanded operational construct and a novel instrument to chart school-relevant language proficiency in pre-adolescent and adolescent learners. *Applied Psycholinguistics, 36*, 1077-1109. doi:10.1017/S014271641400006X

Uccelli, P., & Meneses, A. (2015). Lenguaje académico: construcción y validación de un instrumento para medir habilidades de lenguaje académico en estudiantes de primaria. Invited paper in Special Edited Volume, M. Brisk & P. Proctor (Eds.), *Miríada Hispánica, 10*, 177-202.

Uccelli, P., Phillips Galloway, E., Aguilar, G., & Allen, M. (2016). Academic languages and bilingualism Informes del Observatorio / Observatorio Reports. 022-06/2016EN ISSN: 2373-874X (online) doi: 10.15427/OR022-06/2016EN

Uccelli, P., & Phillips Galloway, E. (2016). Academic language across content areas: Lessons from an innovative assessment and from students' reflections about language. *Journal of Adolescent & Adult Literacy.* DOI: 10.1002/jaal.553

Appendix 1:

CALS-I Tasks Descriptions

(Uccelli & Phillips Galloway, 2016)

Tasks	Skills measured	Sample items	Additional examples
Unpacking dense information: Complex words (selected items from Kieffer, 2009[a]; Kieffer & Lesaux, 2012[b]; adapted from Carlisle, 2000[c]) and complex sentences (selected and adapted items from version 2 of the Test for Reception of Grammar; Bishop, 2003[d])	Skill in breaking down complex words	The administrator reads a morphologically derived word followed by an incomplete sentence, and students are asked to complete the sentence by extracting the base from the derived word (e.g., "*Ethnicity*. The city had many ___ groups.").	*invasion, durability, contribution*
	Skill in understanding complex sentence structures	The administrator reads a sentence, and students are asked to select the picture that corresponds to the target sentence. Four pictures are presented, three of which depict sentences altered by a grammatical element (e.g., "The sheep the girl looks at is running.").	expanded noun phrases, center-embedded clauses
Connecting ideas logically	Skill in understanding school-relevant words that connect ideas	Students are asked to select the missing marker from among four options (e.g., "Kim was sick ___ she stayed home and did not go to school. *otherwise, yet, in contrast, as a result*").	*consequently, nevertheless, in conclusion*
Tracking participants and themes	Skill in tracking referents through a text	Students are asked to match the underlined text with its antecedent by selecting among three options (e.g., "China resisted the move for change. In 1989 students protested to demand changes, but the army opposed these changes. Troops were sent to stop the movement.").	tracking references for concrete participants, events, abstract ideas
Organizing analytic texts	Skill in argumentative text organization	Students are asked to order four to six fragments of a brief essay (introduced by conventional markers; e.g., *in my opinion, one reason, in conclusion*) in order to display a conventional argumentative text structure.	"Some think...," "Others think...," "The first reason...," "The second reason..."
Understanding metalinguistic vocabulary	Skill in understanding words that label or qualify language or thinking moves	The administrator reads two sentences from an informational article followed by a one-sentence reaction from a respondent. Students are then asked to select which word best describes the respondent's reaction from among four options (e.g., *paraphrase, generalization, hypothesis, contradiction*).	*counterclaim, evidence, precise*
Interpreting writers' viewpoints	Skill in interpreting markers that signal a writer's level of certainty about a claim	The administrator reads a "scientist's" claim that includes a stance marker, and students are asked how sure they think the scientist is about the claim made (e.g., "Certainly, the rock is from space."). Students select from among four options to answer the question (e.g., "Is this scientist sure that the rock is from space? yes, maybe yes, maybe no, no").	*impossible, presumably, conclusively*
Recognizing academic register	Skill in identifying more academic versus more colloquial language	Students are asked to select the most academic definition from a set of three definitions of the same familiar word.	*umbrella, clown, debate*

[a]Kieffer, M.J. (2009). *The development of morphological awareness and vocabulary knowledge in adolescent language minority learners and their classmates* (Unpublished doctoral dissertation). Harvard University, Cambridge, MA. [b]Kieffer, M.J., & Lesaux, N.K. (2012). Effects of academic language instruction on relational and syntactic aspects of morphological awareness for sixth graders from linguistically diverse backgrounds. *The Elementary School Journal*, *112*(3), 519–545. [c]Carlisle, J.F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing*, *12*(3), 169–190. [d]Bishop, D.V. (2003). *Test for Reception of Grammar version 2 (TROG-2)*. Oxford, UK: Pearson.

Appendix 2

Table 1: Demographic characteristics CALS-I Study 1 sample.

	n (%)
Gender	
Female	114 (52%)
Male	104 (48%)
SES	
Free/reduced lunch	175 (80%)
No free/reduced lunch	43 (20%)
Ethnicity	
Black/African American	143 (66%)
White	48 (22%)
Latino/Hispanic	15 (7%)
Asian	2 (1%)
American Indian/Alaskan Native	1 (.5%)
Two or more races	9 (4%)
Language Status	
Classified as English Language Learners	39 (18%)
Classified as English proficient	179 (82%)
Language minority students	66 (28%)
Special Education Status	
Classified as SPED	34 (15%)
Not classified as SPED	184 (84%)
Grade	
4^{th}	52 (22%)
5^{th}	55 (23%)
6^{th}	39 (17%)
7^{th}	50 (21%)
8^{th}	39 (17%)
Total	235

Table 2: Demographic characteristics CALS-I Study 2 4th-to-6th grade sample.

	n (%)
Gender	
Female	107 (49%)
Male	111 (51%)
Grade	
4th	78 (36%)
5th	58 (27%)
6th	82 (37%)
SES	
No free/reduced lunch eligibility	77 (35%)
Free/Reduced-price lunch eligible	141 (65%)
4th	46 (59%)
5th	44 (76%)
6th	51 (63%)
Language Status	
Classified as English proficient	109 (50%)
Classified as English Language Learners	109 (50%)
4th	38 (49%)
5th	30 (52%)
6th	41 (50%)
Ethnicity	
Black/African American	65 (30%)
White	43 (19%)
Latino/Hispanic	89 (41%)
Asian	15 (7%)
Two or more races	6 (3%)
Special Education Status	
Classified as SPED	30 (14%)
Total	218

Table 3. Demographic characteristics for the CALS-I Norming Study sample.[3]

[3] Socio-demographic information was not available for all participants.

	n (%)
Gender	
Female	3,338 (50.1%)
Male	3,321 (49.9%)
Grade	
4th	2,563 (36%)
5th	987 (14%)
6th	2,109 (29%)
7th	1,228 (17%)
8th	265 (4%)
SES	
No free/reduced lunch eligibility	1,263 (19%)
Free/Reduced-price lunch eligible	5,382 (81%)
Language Status	
Classified as English proficient	5,825 (88%)
Classified as English Language Learners	822 (12%)
Ethnicity	
Black/African American	2,739 (42%)
White	1,739 (27%)
Latino/Hispanic	1,622 (25%)
Asian	191 (3%)
Native American or Pacific Islander	47 (0.7%)
Two or more races	111 (1.7%)
Other	11 (.6%)
Total	7,152