UCHICAGO Consortium on School Research

TECHNICAL REPORT JANUARY 2022

CPS Gradebook Technical Report

Briana Diaz and John Q. Easton

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

This has been a complex and time-consuming project that we could not have undertaken without the help and encouragement of our many partners at Chicago Public Schools. First, former Chief Education Officer Dr. LaTanya McDade encouraged our broader research agenda and requested this specific analysis. Special thanks to Patrick Gallagher, Director of Student Information Systems, and William Clark, Applications Manager, for providing access to the Gradebook data to Jared Sell, External Research Strategist. Jared patiently walked us through the many tables that we would need and suggested keys to link them together. Sarah Dickson, External Research Manager, supported us throughout the whole process. At the UChicago Consortium, our IT team, led by Richard Blocker, figured out how to transfer and then store such large data files. Todd Rosenskranz, Data Archivist, found a home on our servers where we could securely store and analyze the data.

We learned an enormous amount about Gradebook from Barton Dassinger, Principal of Chavez School. He taught us the many capabilities of Gradebook, answered our many questions, provided feedback on an earlier version of the paper, and continued to encourage us. Quinton Keith of the Network for College Success; Andrew Chipman, Assistant Principal of Thomas Kelly College Preparatory; and Thomas Kelly-Kemple, former Programmer at Benito Juarez High School, also provided much appreciated assistance.

We received very helpful feedback from our colleagues Elaine Allensworth, Alyssa Blanchard, Holly Hart, Jessica Tansey, and David Stevens on the report. We also thank research team member Lauren Sartain, University of North Carolina Chapel Hill, for her advice, former team members Naureen Kheraj and Silvana Freire, and colleagues Vanessa Gutiérrez and David Stevens for their contributions to this research. Shelby Mahaffie completed a very thorough technical review and we thank her for offering helpful comments. Jessica Tansey and Jessica Puller from the UChicago Consortium's communications team carefully edited the report and carried it through to production.

This Gradebook investigation was funded by the Hyman Milgrom Opportunity Fund at the University of Chicago. The Consortium Investor Council funded start-up work on this study and the Hewlett Foundation and an anonymous donor funded our continuing work.

Finally, the UChicago Consortium gratefully acknowledges the Spencer Foundation and the Lewis-Sebring Family Foundation, whose operating grants support the work of the UChicago Consortium, and also appreciates the support from the Consortium Investor Council that funds critical work beyond the initial research: putting the research to work, refreshing the data archive, seeding new studies, and replicating previous studies. Members include Brinson Family Foundation, CME Group Foundation, Crown Family Philanthropies, Lloyd A. Fry Foundation, Joyce Foundation, Lewis-Sebring Family Foundation, McCormick Foundation, McDougal Family Foundation, Polk Bros. Foundation, Spencer Foundation, Steans Family Foundation, Square One Foundation, and The Chicago Public Education Fund.

Cite as: Diaz, B., & Easton, J.Q. (2021). *CPS Gradebook Technical Report.* Chicago, IL: University of Chicago Consortium on School Research.

This report was produced by the UChicago Consortium's publications and communications staff: Jessica Tansey, Communications Manager; and Jessica Puller, Senior Communications Strategist.

Graphic Design: Jeff Hall Design
Photography: John Q. Easton
Editing: Jessica Tansey and Jessica Puller

01.2022/PDF/jh.design@rcn.com

Introduction

In this short report, we explore teachers' use of Gradebook—Chicago Public Schools (CPS)'s electronic grade tracking system—in ninth-grade math classes in CPS in the 2016–17 school year. This technical report concludes one phase of a larger Consortium study of gender differences in CPS ninth-grade math courses. The results of that study will be published separately in summer 2022.

CPS officials requested that we investigate Gradebook usage independently of our larger research question and prepare a short report describing it so that they could understand trends and implications about grading practices. However, because our larger study pertains to grades in ninth-grade math, this report is restricted to that subject and grade level.

About 15 years ago, Gradebook replaced traditional paper gradebooks with a more versatile electronic system to help teachers keep track of students' marks and calculate their grades. Gradebook suggested default grading categories and weights that teachers could alter easily. It also provided parent and student portals, allowing them to monitor student progress. Local school administrators were also able to track patterns and trends in student performance within their own school in Gradebook. They could create a wide variety of reports to look across subject areas or grade levels. (District level administrators also had access to Gradebook, but not authority to change local decisions.)

Although Gradebook is no longer used in CPS, ASPEN, its successor, is very similar and the patterns we observe here likely apply to ASPEN as well. While Gradebook was, and ASPEN is now, ubiquitous in CPS, there is little system-wide written guidance about their use. The one document that we have located is *Professional Grading Standards* and *Grading Practice Guidelines for Chicago Public School Teachers*, developed by a joint committee of CPS and the Chicago Teachers Union (CTU) in 2017.[1] We refer to this CPS-CTU guidelines document several times within this report.

We hope that this short report will prove to be helpful to CPS principals, teachers, and network and central office staff in their reflections and decisions about grading practices in their classrooms and schools.

1 Chicago Public Schools (2017).

Data

To the best of our knowledge, Gradebook data from CPS have never been used before in external research, nor carefully examined and analyzed either internally or externally. Thanks to major efforts from CPS and Consortium data experts, we obtained the 2016–17 school year Gradebook files in the summer of 2020. It took Consortium staff approximately six months to reconstruct and connect five of 13 different "back end" Gradebook tables into a single file that more closely resembled what "front end" users see. (**See Appendix A on p.11 for names of specific files used and how they were connected.**) In terms of data points, this was the largest data file that the UChicago Consortium ever received from CPS.[2]

We extracted both fall and spring semester algebra and geometry classes of ninth-graders in the 2016–17 school year from the complete Gradebook data set. Since CPS charter schools did not report student grades to CPS, the analysis excluded Chicago charter high schools. **Table 1** contains counts of schools, math teachers, math sections, and students observed in the following analyses. We defined math sections as algebra and geometry classes that enrolled ninth-grade students. For most of the analyses, we used the section as our unit of analysis instead of teacher, because a single teacher may use different categories and weightings in different sections.

TABLE 1

Description of Analytic Sample SY 2016–17

Sample	Records
Students	17,056
Teachers	366
Schools	86
Sections	1,804

Note: We defined math sections as algebra and geometry classes that enrolled ninth-grade students.

Key Terms

Throughout this report, we use multiple terms that have specific meanings that are not intuitive and require explanation. Some of these terms come directly from Gradebook, and we have created others from analyzing the raw data in Gradebook. **Appendix B** provides more detail, but the most important to keep in mind are the following:

- **Category Titles.** Gradebook provides six default category titles: **1)** assignments, **2)** homework, **3)** class participation, **4)** quizzes, **5)** exams, and **6)** projects. Teachers can alter these titles, using category titles they find more appropriate, such as "quizzes/tests." Teachers can also add up to three additional categories and provide the titles themselves. Categories are like "buckets" that contain multiple tasks (called "assignments" in Gradebook).
- **Category Family.** This term refers to our own coding of the 709 unique category titles that we observed in the data files into larger buckets called "families." We did this to create a more parsimonious set of categories to make our analysis more interpretable. These are presented in **Table 2**.
- **Task.** We use term "task" to refer to the multiple pieces of work within each category title or family. We chose to call these tasks to avoid having two separate meanings of Assignments (a category title, in addition to the units contained in the bucket). An example of a task is "Problems 1-5, page 38."

2 Gradebook data use at the UChicago Consortium is governed by a legally valid data sharing agreement with CPS. Per this agreement, the UChicago Consortium holds no Personally Identifiable Information, such as student names.

Findings

1. Most teachers customized category titles.

Very few ninth-grade math sections used the default category titles as suggested in Gradebook. About 2 percent of the sections in this sample used exclusively default category titles—most often using two, three, or four of the default categories. In 98 percent of the sections, teachers either altered category titles or created new ones.

There were 709 unique category titles across the 1,804 sections in this analysis. To reduce often-repetitive and similar category titles we coded them into six "category families" (**see Appendix B on p.13 for a description of our coding process and checking**). **Table 2** below shows the six category families and their relative prevalence.

We coded all 709 category titles into these six category families; they are mutually exclusive. We coded 43 percent of the category titles into the assessments category family, and 27 percent into the assignment category family, etc. Note that the percentages in the table sum to 100 percent (without rounding).

Coding the category titles into families led us to four additional category title findings:

- **We could not differentiate between formative and summative assessments.** The Gradebook default categories separate tests and quizzes, and the CPS/CTU agreement on Gradebook usage requires the separation of summative and formative tests.[3] Yet, in our coding, we often were unable to differentiate between tests and quizzes and between summative and formative. For example, teachers created

TABLE 2

Category Families as Coded from Multiple Teacher-Assigned Category Titles

Category Family	Definition	Percentage of Category Titles Coded into Each Family
Assessments	Categories that evaluate a student's knowledge of a topic. Quizzes, exams, etc.	43%
Assignments	Categories applied to work given in the day-to-day course of a class. Classwork, homework, bell ringers, etc.	27%
Behavior	Categories that evaluate a student's actions in the class. Behavior, class participation, etc.	10%
Mastery-Based	Categories that follow the Common Core State Standards or general mathematics principles. Number sense, inequalities, etc.	7%
Weekly	Categories organized by week administered. Week 5–10, Week 30, etc. Typically the last week's data applied to final grade.	3%
Other/Unclassifiable	Any category unable to be grouped into one of the category families above.	11%

Note: Aggregate values based on unique category records. N=709

3 Chicago Public Schools (2017).

category titles such as "quizzes and tests," "exams/quizzes," "assessments (summative/formative)" and "quizzes, exams, projects, and other assessments." As we continue to refine our work with Gradebook, we hope to develop stronger rubrics to differentiate between summative and formative assessments, as we believe that there are meaningful differences in how the two types of assessments are used.[4] We return to this subject in the final section of this report.

- **We could not differentiate assignments into classwork and homework categories.** They were often combined in teachers' category titles, such as "classwork and assignments," and "classwork/homework assignments." This category family contains 27 percent of all category titles.
- **The "mastery-based" category family was used relatively commonly.** Seven percent of unique category titles reflected specific learning standards being covered in class. Examples include: "real life diii/iv/v. apply and justify," "communicate (ciii/iv/v) move and organize," or more simply, "systems and equations." This likely reflects that several CPS high schools have adopted standards- or competency-based grading, where grades are determined by students' performance on assessments or other demonstrations of content mastery. The "weekly" category family contains category titles such as "Weeks 1-5" and "Weeks 5-10." Presumably, these refer to content covered during the specified period and may reflect grades calculated by teachers outside of Gradebook for the named grading period.
- **We were unable to code 11 percent of category titles because we did not understand their meanings, or because their meanings crossed category families.** For example, "projects/labs/essays," "direct learning," and "quizzes, classwork, projects" were categories placed into the other/unclassifiable category family.

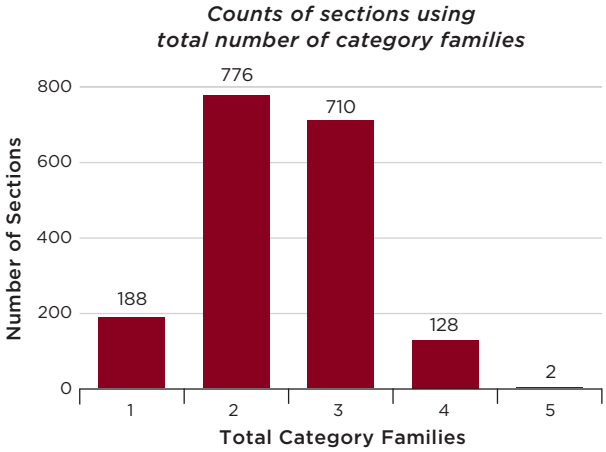
2. Most teachers used two or three grading category families.

Gradebook permits customization of category titles to suit teachers' pedagogical preferences. The average number of category families used in ninth-grade math sections was 2.43; most teachers used two or three grading category families (**see Figure 1**). About 80 percent of sections used two or three grading categories families. Again, we see that few sections used the six default grading category titles. This appears to reflect teachers' grading preferences and teaching and grading practices.

Note that teachers used a larger number of unique category *titles*—3.78 on average. The average number is reduced because we collapsed many similar sounding titles into the larger *family*. No doubt individual teachers could meaningfully differentiate similar sounding category titles, but we could not. **See Appendix D on p.15 for more information.**

FIGURE 1

Teachers Typically Used Two or Three Category Families



Note: Category families sometimes contained multiple category titles; see Key Terms on p.2 for details. N = 1,804 sections

4 We consider formative assessments to be assessments given to gauge how well students are absorbing information as instruction is happening, in contrast to summative assessments, which are administered at the end of the instructional period.

3. In sections that used mastery-based and weekly category families, almost of all the final grade was determined by those category families. In other sections, assessments had the highest weighting.

Table 3 notes the relative usage of our five category families. The median percentile column shows the weight for grades in each category family in the middle of the distribution—one-half of sections had higher weightings and one-half had lower weightings. The 25th percentile is the point in the distribution where 25 percent of section had lower weightings and 75 percent had higher weightings. In the same vein, the 75th percentile is the point where 25 percent of sections had higher weightings and 75 percent had lower weightings.

In sections where teachers used mastery-based or weekly category families, nearly all the final grade was determined by tasks within these respective category families. This was likely due to mastery-based or weekly grading schemes being a different framework from which to approach grading—and if a teacher used a non-traditional scheme, they would consistently use that non-traditional scheme.

In sections that did not use mastery-based or weekly grading, assessments had the highest weighting in determining final grades, followed by assignments, and then behavior. As noted above, often we could not differentiate tests from quizzes, and summative assessments from formative assessments. Because of this difficulty, our median assessment category family weighting of 60.0 percent exceeded the CPS/CTU recommendation that no single category should exceed 50 percent of the total weighting.[5] In future work, we hope to develop more precise rubrics to differentiate formative assessments from summative assessments.

TABLE 3

Typical Final Grade Weights, by Category Family

Category Family	Median (50th Percentile Weight)	25th Percentile Weight	75th Percentile Weight
Assessments	60.0	50.0	70.0
Assignments	30.0	20.0	40.0
Behavior	10.0	10.0	11.1
Other	25.0	10.5	50.0
Mastery	80.0	70.0	85.0
Weekly	100.0	100.0	100.0

Note: Aggregate values based on unique section/category family/student records. N=86,533

5 Chicago Public Schools (2017).

4. On average, all category families included many tasks; the exception was in the weekly category family, where typically only four tasks contributed to final grades.

Table 4 provides evidence that students completed multiple tasks within each category family, so that no single task played an outsized weight in determining the category scores and thus final scores. On average, there were about 11 tasks per semester in the assessments category family—approximately one assessment task every three weeks over the course of a full semester. There were about 28 tasks on average in the assignments category family.

The median percentile column shows number of tasks recorded in each category family in the middle of the distribution—one-half of sections had more tasks and one-half had fewer. The 25th percentile is the point in the distribution where 25 percent of section had fewer tasks and 75 percent had more. In the same vein, the 75th percentile is the point where 25 percent of sections had more tasks and 75 percent had fewer.

Teachers who used mastery-based grading assigned around 21 tasks for each semester in that category family, on average. Teachers who used weekly grading schemes tended to consistently use four tasks for final semester grades. The weekly category family is hard to interpret given that the category titles in this family varied greatly. For example, some were titled "weeks 1-5," or "week 6" or "weeks 4-7."

TABLE 4

Typical Number of Tasks by Category Family

Category Family	Median (50th Percentile Task Count)	25th Percentile Task Count	75th Percentile Task Count
Assessments	11.0	8.0	14.0
Assignments	28.0	16.0	46.0
Behavior	7.0	2.0	17.0
Other	14.0	6.0	25.0
Mastery	21.0	14.0	30.0
Weekly	4.0	4.0	4.0

Note: Aggregate values based on unique section/category family/student records. N=86,533

5. Sixty-five percent of sections used the Gradebook default weighting algorithm.

Final grades were affected by two types of weightings. The first was the category (title or family) weighting. Gradebook assigns default weights to the six default categories, as follows: assignments, 40 percent; homework, 20 percent; class participation, 10 percent; quizzes, 10 percent; exams, 10 percent; and projects, 10 percent.

The second type of weighting occurs at the task level. Tasks within grading category titles may also be weighted differently from each other by assigning different total points possible to individual tasks. For example, one quiz may be worth 100 points, another 60, and a third 50.

These weightings may be combined in two different ways:

- "Total Points Logic ON." Or
- "Total Points Logic OFF." [6]

The default setting for Total Points Logic is "ON." This means that the task scores are aggregated to the category title score by dividing total points earned by total points possible (**see Table 5**). For example, if a teacher gave three tests (or tasks) with different points possible, the tests with more possible points would be more influential in determining the final category scores. **Table 5** provides a specific example.

Using Total Points Logic (ON) this student's category score would be 180 divided by 210=85.7 percent.

Total Points Logic (OFF) treats each test as if it had equal weight to the others regardless of differences in points possible. So, in the example shared in **Table 5**, it would average 80 percent, 92 percent, and 90 percent to result in a category score of 87.2 percent. Tests #1, #2, and #3 had equal influence in determining the final score, even though they had different possible points.

The category title weights (for example, exams being 10 percent of final grade) are then applied to these category scores to calculate the final grade. About 65 percent of sections in this sample used the default (Total Points ON) and 35 percent used Total Points Logic OFF. There is a subtle difference between the two, and our anecdotal reports suggest that it is not well understood by most administrators and teachers.

The example above shows a relatively small difference between the Total Points Logic options, but it is very possible that a student could earn a different final grade depending on whether the Total Points Logic option is either ON or OFF.

TABLE 5

Possible Category Score Calculation

Test	Points Earned	Points Possible	Percent Correct
#1	80.0	100.0	80.0
#2	55.0	60.0	91.7
#3	45.0	50.0	90.0
Total	180.0	210.0	—

6 The ASPEN system now used by CPS provides four optional weighting systems.

6. Students on the cusp between two grades were much more likely to receive the higher grade than the lower grade.

Our final finding is not related to category titles and category families, but rather, students' final grades. Gradebook afforded us the ability to observe student's final grade values, not just their final letter grade. Plotting all final grade values from our sample allowed us to ascertain grading patterns throughout CPS.

Figure 2 shows the distribution of final course scores from fall and spring semesters, combined. There were very few students on the left side of the grade distribution (below 60) and an even smaller number on the right side with scores exceeding 100. This is likely due to extra points and additional tasks.

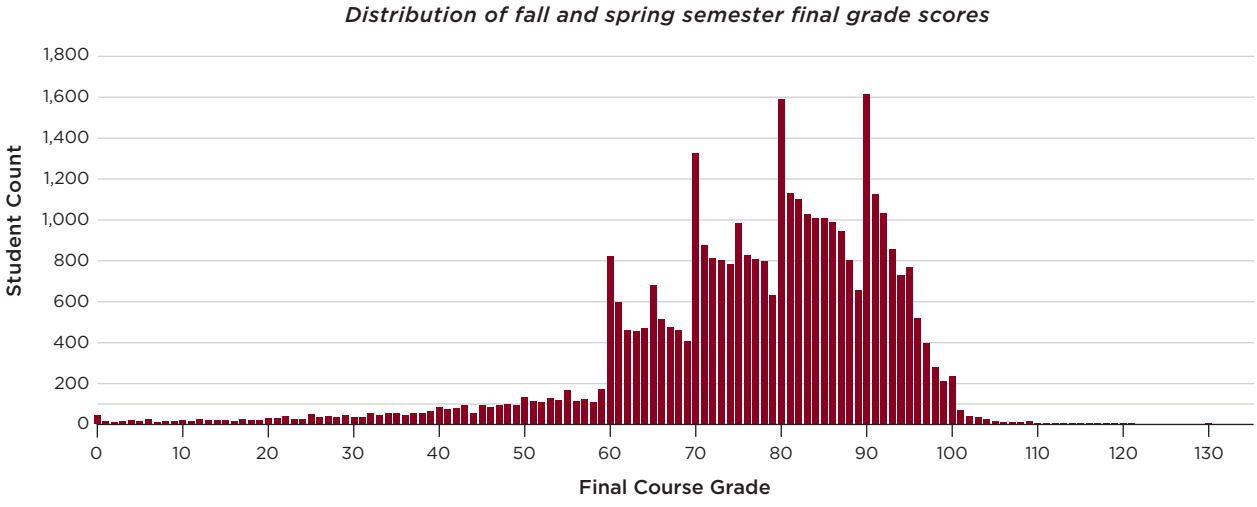
In examining this figure, one is immediately drawn to the four "peaks" where there are very few students immediately to the left (those with lower scores) and a sharp decline in the number of students immediately to the right (those with higher scores).

These peaks occur at 60, 70, 80, and 90 points—the cutoffs between F and D, D and C, C and B, and B and A, respectively. Many students earned the fewest possible points that would earn a higher grade. In fact, the most common point score for a given grade was the minimum for that grade. Given the available data, we cannot determine how this is happening, but many students "on the border" were able to move up into the next higher letter grade.

There are many possible explanations—speculations, given our lack of hard data—for this frequent occurrence. One possibility is that scores on the cusp were "rounded up." Or teachers may be providing students with extra opportunities to earn credits to move to a higher grade. Students, eager to pass or earn a higher grade, may be requesting extra credit tasks. Or teachers and students are working together to avoid failures and to earn higher grades. These practices would be in keeping with the CPS/CTU agreement that states: "To encourage student success, students shall be allowed the opportunity to recover and improve."[7]

FIGURE 2

Students on Cusp Between Two Grades Were More Likely to Receive Higher Grade



Note: Fall and spring final grades. N=32,234

7 Chicago Public Schools (2017).

Implications

In reviewing and analyzing Gradebook data, we were able to learn much about teacher grading practices in CPS. Before analyzing the data, we expected that most teachers were using the default, suggested CPS/CTU categories, a point that the data did not bear out. Most teachers heavily edited their Gradebook categories—changing the number of categories used, the names of those categories, and the weight of the category toward a student's final grade. Ten percent of teachers rejected grading based on type of assignments altogether, instead opting for grading schemes based on math standards or when tasks were assigned.

After coding individual category titles into category families, we were able to gain insights into teachers' grading practices and decisions that affected final grades. We found that while category titles in the assignment category family had the greatest number of tasks counting toward the final grade, assessment family categories typically had the most weight toward a student's final grade. Additionally, we found that teachers in our sample typically did not make clear distinctions between formative and summative assessment categories in Gradebook, and they often created category titles that encompassed both types of assessments.

Finally, we were able to observe that teachers seem to be providing opportunities for students to improve their grades throughout the semester. A relatively small proportion of students had final grade scores below the 60 percent value, and many students received the higher of two letter grades when they were on the cusp between grades at the end of the semester.

This brief report analysis raises several key issues for further discussion:

- **This analysis, together with our anecdotal conversations with current and former CPS educators, suggest that there is little formal guidance for the use of Gradebook and its successor, ASPEN.** Although there are pockets of expertise in CPS about the use of electronic grading systems, they are not widespread. This is perhaps one reason why we find such great variability in Gradebook usage in a single subject at one grade level.
- **Users may be helped by greater clarification around grading practices between formative and summative assessments.** In coding teacher-developed category titles associated with assessments, often we could not differentiate summative from formative assessments, and tests from quizzes. This strikes us as a critical issue since the two are different and serve different purposes. There is also the question of whether formative tests should even count in grades, as they can be intended as a learning tool only. This will become even more important as CPS begins to implement Skyline, which contains a wider range of assessment types. Teachers will need to decide whether and how to include different assessment types in ASPEN. They may also need guidance in making these decisions.
- **Teachers using mastery-based grading may want to create their own best-practices community.** The Gradebook data reveal great variability within our mastery-based category family, suggesting school-to-school differences in implementation. We often saw only one or two teachers at a school site using mastery-based grading, while at other sites a greater proportion of teachers used mastery-based grading. Given the growing popularity of standards-based grading, schools participating may learn best-practices from each other by discussing how they use Gradebook (now ASPEN).
- **The weighting options and "Total Points Logic" are difficult to understand and to explain.** Again, this topic deserves greater attention so that teachers understand the implications of their choices. This is even more important in ASPEN than in Gradebook, given the greater number of options available.

- **Further investigation of "cusp" grade patterns would provide further insights into grading practices.** **Figure 2 on p.8** provides interesting evidence about grading practices. Do the four peaks reflect evidence that teachers are providing and students are requesting more opportunities for student success, as encouraged in the CPS/CTU Grading Standards? This topic deserves more conversation about how this is occurring and how more students can reach CPS's "Bs or better" goal.
- **Now is the time to update the CPS/CTU Grading Standards publication.** Given the variability in grading, the rising popularity of mastery-based grading, and the coming introduction of Skyline, now may be a good time to revisit and revise Grading Standards and make it widely available through professional development experiences. This may provide the opportunity for more in-depth discussions on various types of assessments and their role in determining grades.
- **Grading patterns in subjects beyond ninth-grade math should be examined.** This investigation is restricted to ninth-grade math classes from several years ago. Now that ASPEN is well established and well known, CPS may want to expand this exploration to other grade levels and subjects, especially those where there may be concerns about student progress or levels of success.

Final Note

Gradebook provided teachers a great deal of latitude and flexibility in its use. We have noted several times how much variability there was in grading practices across math sections. We are not suggesting that greater consistency in grading practices is necessarily desirable. But it is likely that greater communication among administrators and teachers about Gradebook (now ASPEN) use could lead to common learnings. These could result in good practices that improve use, are easier for teachers, and facilitate communication among teachers, students, administrators, and parents.

References

Chicago Public Schools. (2017)
Processional grading standards and grading practices guidelines for Chicago Public Schools teachers. Retrieved from https://www.cps.edu/globalassets/cps-pages/staff/employee-engagement/professional_grading_standards.pdf

Appendix A

Gradebook Database Tables

We used the records and file names in **Table A.1** to re-create final student grades; **Figure A.1** displays the connections across data sources.

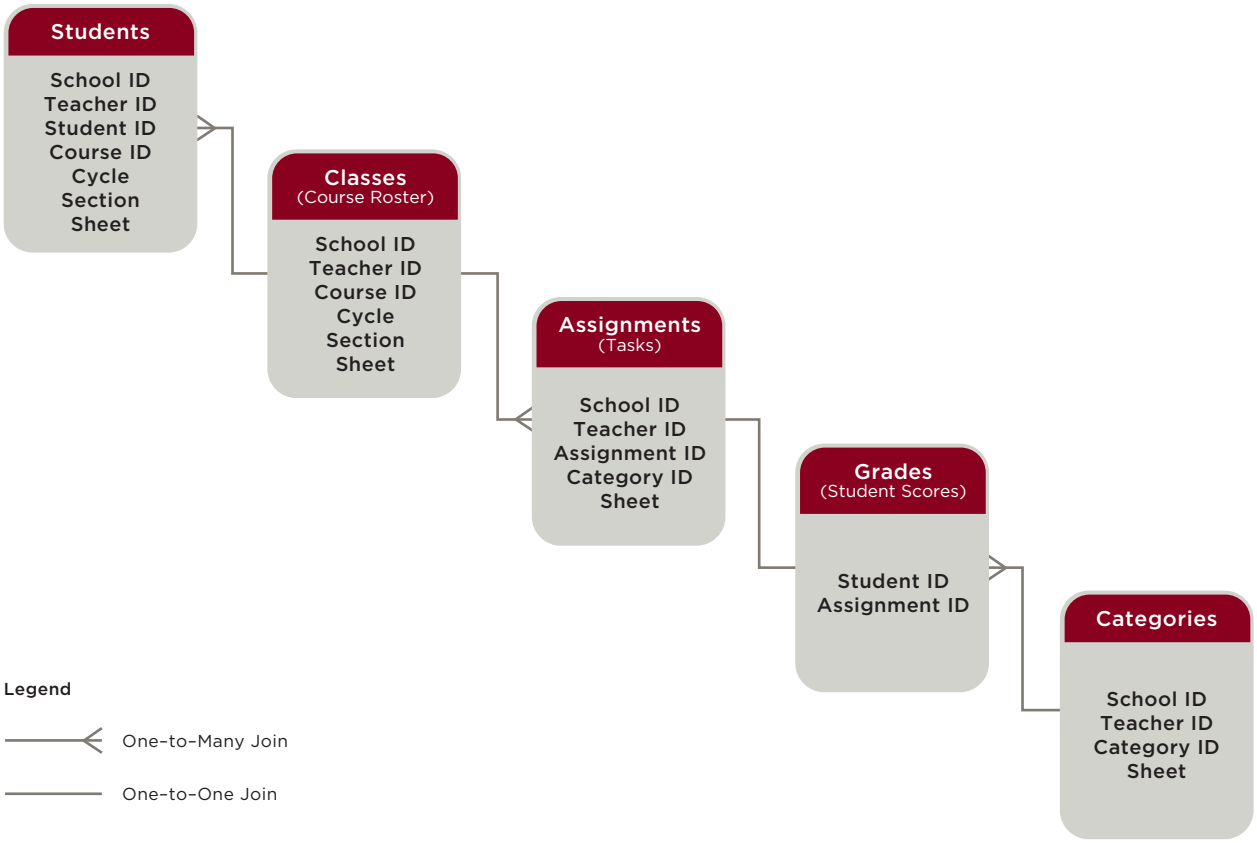
TABLE A.1

List of Gradebook Tables and Their Connections

Table	File Name	Records
Course Overrides	gbcourseoverrides	1,684
Category Groups	gbcatgroups	36,550
SS Assignment Links	ssassignmentlinks	173,973
Teacher ID Map	imsteacheridmap	21,799
Categories	gbcategories	3,479,287
Classes	gbclasses	714,024
Assignments	gbassignments	7,965,950
Grades	gbgrades	159,209,418
Comment Code	gbcommentcd	47,443
Students	gbstudents	14,931,021
Course File	gbcoursefile	1,161
Student Overrides	gbstudentoverrides	1,420
Quarter 1 Report Card	reportcards_q1_2016_17_hs	546,440
Quarter 2 Report Card	reportcards_q2_2016_17_hs	566,735
Quarter 3 Report Card	reportcards_q3_2016_17_hs	529,411
Quarter 4 Report Card	reportcards_q4_2016_17_hs	434,647

FIGURE A.1

Connections for Analytic File



Note: Data included here represent the data in tables shared with Consortium researchers; CPS data may contain additional variables.

Appendix B

Definitions and Examples of Terms Used throughout This Report

We use multiple terms throughout this report that have specific meanings that require explanation. Some of these terms come directly from Gradebook, and we have created others from analyzing the raw data in Gradebook. **Table B.1** provides definitions and examples for our key terms.

TABLE B.1

Key Terms

Term	As Defined By	Definition	Example
Default Category Title	Gradebook	Title or name given to a gradebook category. Can be either one of six default titles, or a teacher-created title.	Assignments Homework Class Participation Quizzes Exams Projects Unnamed Category 1 Unnamed Category 2 Unnamed Category 3
Modified Category Title	Gradebook	Category names modified or created by teachers.	"In class learning" "Ticket to leave" "Formative: bell ringer and group work" "Math standard 3.2"
Category Family	Consortium Researchers	High-level grouping of category titles. Categories were *coded* into category families by Consortium researchers	Assessments Assignments Behavior Mastery-Based Weekly Other/Unclassifiable
Task	Consortium Researchers	Work assigned under each category. Gradebook refers to these as "assignments." To reduce confusion with the assignment category family, we refer to these as tasks. There are usually multiple tasks within each category family.	"Slope and graphing review" "Quiz 3- Word problems" "Math Binder" "Graphing lines in standard form hw" "Problem Set #5 – Parallel/ perpendicular lines"

Appendix C

Category Family Coding

Initially, our team intended to use the default Gradebook category titles to perform this analysis. Once we received the Gradebook data from CPS, it became clear that few teachers (40 percent) used at least one of the default category titles with its default weight, and that to perform any meaningful analysis with categories we would need to create a taxonomy of category families. The initial set of math courses taken by first-time ninth-graders contained 1,100 unique category titles in addition to the six default titles. After we limited the set to algebra and geometry courses, 709 unique category titles remained, but all coding was done on the original set. We limited the analysis to the two most common ninth-grade math courses, Algebra and Geometry, after initial coding was finished.

First, we created a category coding scheme of three major category families: Assessments, assignments, and behavior. Where we observed teachers using math standards for category titles, we placed them into the mastery-based category. Similarly, where we observed teachers listing weeks of the course as category titles, we placed the category into the weekly family. All other category names that we could not group into one of the category families above were placed into "other/unclassifiable."

Quality Assurance and Validation

The initial category coding was performed by Briana Diaz. After the initial round of category coding, John Easton and Silvana Friere each performed validation checks on approximately one-fourth to one-third of the categories, and their feedback was incorporated into the final coding rules. The data in this report reflects the data from the first round of validation. In summer 2021, our team worked with former K-12 teachers to perform another round of validation; the results of that work are forthcoming.

The current coding scheme is a work in progress and may change in the future. For example, we are attempting to split the assessments category family into two categories: formative and summative assessments. However, it remains to be determined if this dataset can support another layer of disaggregation.

Appendix D

Additional Tables

This appendix contains additional reference tables. **Table D.1** displays the number of sections by the total number of category titles used. **Table D.2** shows the pre-set default category titles and weights.

TABLE D.1

Total Category Titles by Sections

Total Category Titles Used	Number of Sections	Percent of Total Sections
1	82	4.5%
2	148	8.2%
3	514	28.5%
4	512	28.4%
5	334	18.5%
6	167	9.3%
7	37	2.1%
8	7	0.4%
9	3	0.2%

Note: Aggregate values based on unique section/category title/student records
N= 86,533

TABLE D.2

Gradebook Category Default Titles and Weights

Default Category Title Name	Default Category Title Weight
Assignments	40%
Homework	20%
Class Participation	10%
Quizzes	10%
Exams	10%
Projects	10%

ABOUT THE AUTHORS

BRIANA DIAZ is a Research Analyst at the UChicago Consortium. Prior to joining the UChicago Consortium, Briana worked in New Orleans public education for nine years, serving as a middle- and high-school science teacher, school district data analyst and database architect, and policy researcher. Her research interests include educational accountability systems, district budgets and finance, and school leadership.

JOHN Q. EASTON is Senior Fellow at the UChicago Consortium. Until 2018 he served as Vice President, Programs at the Spencer Foundation in Chicago. At Spencer, he developed and led a grant program for research-practitioner partnerships. From June of 2009 through August 2014, he was Director of the Institute of Education Sciences in the U.S. Department of Education. Prior to his government service, Easton was Executive Director of the UChicago Consortium. He has been affiliated with the UChicago Consortium since its inception in 1990 and became its Deputy Director in 1997 and Executive Director in 2002. John served a term on the National Assessment Governing Board, which sets policies for the National Assessment of Educational Progress (NAEP). He is chair of the Illinois Workforce and Education Research Collaborative Advisory Board; a member of the Illinois Employment Security Advisory Board, Educational Testing Service External Research Panel, and the Chicago Public Schools' School Quality Report Card 3.0 Advisory Board.

UCHICAGO Consortium on School Research

Steering Committee

PAIGE PONDER
Co-Chair
One Million Degrees

REBECCA VONDERLACK-NAVARRO
Co-Chair
Latino Policy Forum

Institutional Members

SARAH DICKSON
Chicago Public Schools

BRENDA DIXON
Illinois State Board of Education

BOGDANA CHKOUMBOVA
Chicago Public Schools

TROY LARAVIERE
Chicago Principals and Administrators Association

JESSE SHARKEY
Chicago Teachers Union

MAURICE SWINNEY
Chicago Public Schools

Individual Members

NANCY CHAVEZ
OneGoal

JAHMAL COLE
My Block, My Hood, My City

ACASIA WILSON FEINBERG
The Cleveland Avenue Foundation for Education

VERNEE GREEN
Mikva Challenge

MEGAN HOUGARD
Chicago Public Schools

GREG JONES
The Academy Group

PRANAV KOTHARI
Revolution Impact, LLC

AMANDA LEWIS
University of Illinois at Chicago

RITO MARTINEZ
Rito Martinez Consulting LLC

SHAZIA MILLER
NORC at the University of Chicago

KAFI MORAGNE-PATTERSON
UChicago Office of Civic Engagement

CRISTINA PACIONE-ZAYAS
Erikson Institute

LES PLEWA
William H. Taft High School

BEATRIZ PONCE DE LEÒN
Illinois Department of Human Services

CRISTINA SALGADO
City Bureau

ELLEN SCHUMER
COFI

PAM WITMER
Golden Apple Foundation

JOHN ZEIGLER
DePaul University

UCHICAGO Consortium on School Research

1313 East 60th Street
Chicago, Illinois 60637

T 773.702.3364
F 773.702.2010

@UChiConsortium
consortium.uchicago.edu

OUR MISSION With the goal of supporting stronger and more equitable educational outcomes for students, the UChicago Consortium conducts research of high technical quality that informs and assesses policy and practice in the Chicago Public Schools. We seek to expand communication among researchers, policymakers, practitioners, families, and communities as we support the search for solutions to the challenge of transforming schools. We encourage the use of research in policy action and practice but do not advocate for particular policies or programs. Rather, we help to build capacity for systemic school improvement by identifying what matters most for student success, creating critical indicators to chart progress, and conducting theory-driven evaluation to identify how programs and policies are working.

THE UNIVERSITY OF CHICAGO | UEI URBAN EDUCATION INSTITUTE