

Abstract Title Page
Not included in page count.

Title: On the Appropriateness of Surveying Students in 4th and 5th Grades

Authors and Affiliations:

Stuart Luppescu, UChicago Consortium on School Research

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

The UChicago Consortium on School Research has been surveying students, teachers, and principals in the Chicago Public Schools (CPS) since 1991 using what is now known as the *5Essentials* surveys. The developers categorized and grouped survey measures into what they refer to as “The Five Essentials for School Improvement.” The value of these surveys is that they measure school organizational quality that predicts student academic improvement, which has been demonstrated convincingly (e.g., Bryk, et al. 2010). What used to be just a CPS instrument has since been adopted as a school improvement vehicle in several school districts across the United States. For example, the *5Essentials* has been used in Minneapolis, Detroit, Florida, among other places. Since 2013, it has been administered across the state of Illinois.

While the validity of the *5Essentials* in contexts outside of Chicago has generally been confirmed (e.g., Klugman, et al. 2015), there have been some challenges involved in using the surveys beyond the bounds of Chicago. Each of the five essentials is measured using 4 or 5 “measures”—or sets of items that measure the strength of belief about an individual construct. Some of the measures come from teacher responses while others come from student responses (see Table 1 for the list of measures and essentials). These measures are aggregated up to the “essential” level. The Supportive Environment essential is an aggregate of five student measures. The Ambitious Instruction essential is composed of one teacher measure and four student measures. Therefore, in order for schools to receive information about these two essentials, student responses are very important. However, historically, students in grades lower than 6 have not been surveyed using the *5Essentials* due to concerns about the ability of younger students to accurately report on conditions in their schools. This did not present problems when the surveys were only administered in CPS, as most elementary schools in Chicago serve students through grade 8; therefore, there are adequate numbers of students in grades 6-8 whose data could be used to calculate these essential scores. But in the rest of the state, or in other states, there are many elementary schools that only serve students in grades K-5. Schools with such a grade structure do not receive reported data for the essentials that rely on student measures. For those schools, the values for the Supportive Environment and Ambitious Instruction essentials are missing.

To address the concerns raised by school districts outside of Chicago, the current author sought to test whether the expansion of the student survey to students in grades 4 and 5 was feasible. There are number of valid reasons why there may be concerns about surveying students this young, including:

1. Younger students could have trouble reading and understanding survey question text because of lower levels of literacy
2. Younger students could have less ability to accurately assess their observations of the context around them
3. Younger students could be less able to recognize and process their emotional reactions to these observations

Indeed, Levenstein and Luppescu (2015) found evidence for the relationship between grade level and survey measure person fit, comparing students in the lower-middle grades and the upper-middle grades. Al-Tayyip et al. (2002) found that lower levels of literacy interfered with respondents' being able to provide accurate survey data. The author takes this idea further, applying it to an investigation of the psychometric properties of survey measures obtained from students in grades 4 and 5.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

In order to avoid reporting back to K-5 schools on only three of the five essentials, this study tests whether students in grades 4 and 5 can provide reliable, valid survey responses on the *5Essentials*.

Setting:

Description of the research location.

Students in a large, urban school district in the Midwest during the 2014-2015 school year.

Population / Participants / Subjects:

Description of the participants in the study: who, how many, key features, or characteristics.

In the district of focus, the *5Essentials* survey was administered to more than 72,000 students in grades 6-8 and 36,000 students in grades 4 and 5, for comparison. The overall survey response rate for students in grades 6-12 was 79.5%. The younger students' data were not used in the School Quality Reports¹, so the response rates were only about 50%. The results for the younger students were used only for research purposes and not reported back to the schools. The students were administered the items that were used to make the following measures: English Instruction (engl), Math Instruction (math), Peer Support for Academic Work (acno), Academic Press (pres), Safety (safe), Student-teacher Trust (trts), and Academic Personalism (perc). The items that go into each of the measures appear in Tables 2-8 in Appendix B.

Research Design:

Description of the research design.

We made the measures listed in the previous section using Rasch measurement, separately for the two groups: 4th and 5th graders, and 6th through 8th graders. We compared the individual-level reliability, school-level reliability and intra-class correlation for each measure for each group. In addition, we compared the proportion of students in different levels of Rasch person fit for each group.

¹ Schools receive school reports that provide data obtained from the surveys. Schools only receive these reports if they have a response rate greater than 50% for teachers and for students in surveyed grades. The lack of receiving data from the 4th and 5th graders may have reduced the incentives for schools to encourage those students to take the surveys.

Data Collection and Analysis:

Description of the methods for collecting and analyzing data.

Survey response data were collected as part of a district-wide administration of the *5Essentials*. Measures were produced using Rasch analysis. Person fit (aggregate information-weighted residuals) was calculated for each person measure. In addition, person-lexile scores as measures of the students' reading ability were obtained from school records. Similarly, the lexile reading difficulty of the survey items in a particular measure was obtained using the lexile online tool. Table 9 in Appendix B gives the average lexile scores of students by grade. Table 10 shows the lexile reading levels of the item text for the seven measures considered in this study.

Findings / Results:

Description of the main findings with specific details.

Individual-level reliabilities: Reliabilities for 4th and 5th grade students were somewhat lower than those of students in grades 6 through 8, but not appreciably so. The biggest difference was for the English Instruction measure. For the younger students the reliability was 0.58, while for the other students it was 0.76. See Figure 1 in Appendix B.

School-level reliabilities: School-level reliabilities were obtained from 3-level HLMs, with a measurement model at level 1, individuals at level 2, and schools at level 3. Similar to the results of the individual-level reliabilities, the school-level reliabilities for students in grades 4 and 5 were somewhat lower than for other students. See Figure 2 in Appendix B.

Intra-class correlations: For all the measures but two, the intra-class correlations (the proportion of variance that is between schools) for the lower-grade students was slightly higher than for other students. For the two measures for which the ICCs were not higher, they were only marginally lower. See Figure 3 in Appendix B.

Fit statistics: To examine the relationship between student grade level and person fit, we categorized the fit statistics into 12 ranges. The fit statistic is on a ratio scale ranging from 0 to infinity. It represents the ratio of observed variation in responses to expected variation in responses, so the expected value is 1.0. We take 1.3 (30% more variation than expected) as evidence of excessive random noise, and is labeled "misfit". Less than the expected amount of variation in the responses (fit statistic < 1/1.3) is labeled as "overfit". The fit categories are calculated using cut points of (0, 1.3⁻⁵, 1.3⁻⁴, ..., 1.0, 1.3, 1.3², ..., 1.3⁵, infinity). The range between 1/1.3 and 1.3 is considered not abnormal. Figures 4 through 10 show the proportion of students in each of the fit categories by grade level for each measure. For example, in Figure 4, fit categories for the Peer Support for Academic Work measure, there are slightly more 4th and 5th graders that misfit, indicating that some of them may have had trouble reading and/or understanding and responding to the items in this measure, but there were far more overfitting students in grades 6 to 8, indicating, perhaps, disengagement and lack of cooperation. Figure 5 shows a somewhat unexpected pattern. The readability level of the text for the measure, English Instruction, as we see in Table 2, is at a 6th to 7th grade level, so we would expect more misfit in this measure for the younger students. In fact, we notice the opposite pattern.

Conclusions:

Description of conclusions, recommendations, and limitations based on findings.

To the surprise of this researcher, the results of administering surveys intended to be used on students in grades 6 to 12, to students in grades 4 and 5 did not show serious evidence of problems in the measurement. The reliabilities of the measures, at the individual as well as at the school level, were slightly lower than for the older students, they were still in the acceptable range, and in no way were they so low that precluded their use. And while there was somewhat more misfit in the measures of the younger students, the more striking finding is the larger amount of overfit (indicating straightlining or some other form of non-cooperation) in older students.

The upshot of having slightly less reliable measures is that the standard errors of measurement will be a bit larger. In addition, we can inflate the standard errors of misfitting students to account for the lower confidence we have in the measures of those students. We routinely use the inverse of the standard errors as weights in statistical analyses to adjust for different degrees of precision in each student's measures. Students whose measures are less precise will then be down-weighted in analyses. So, while the psychometric properties of measures from students in grades 4 and 5 are somewhat less salubrious than students in higher grades, they are not so unfavorable as to preclude measurement at all.

This study came about as a result of recent expansion of the use of the *5Essentials* surveys in areas, locations, and with populations previously unsurveyed. While once only given to students and teachers (and occasionally principals) in Chicago, the surveys are now given around the state of Illinois and in other locations across the United States. Given the needs of schools that only serve students through grad 5, it is natural that the surveys should be extended to students in grades 4 and 5 to fill the gaps in measurement if it can be psychometrically justified. The outcome of this study is that the State of Illinois, and other locations around the country will be able to produce more complete reports for their schools, with less missing data, without fear of disruptive measurement problems.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Al-Tayyib, A., Rogers, S.M., Gribble, J.N., Villarroel, M., & Turner, C.F. (2002). Effect of low medical literacy on health survey measurements. *American Journal of Public Health, 92 (9)*, 1478–1480.

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: The University of Chicago Press.

Klugman, J., Gordon, M.F., Sebring, P.B., & Sporte, S. E. (2015). *A First Look at the 5Essentials in Illinois Schools*. Chicago, IL: UChicago Consortium on School Research.

Levenstein, R. M. and Luppescu, S. (2014). *Student reading ability and survey measure fit*. Paper presented at Annual Conference, Midwest Association for Public Opinion Research, Chicago, IL.

Appendix B. Tables and Figures

Not included in page count.

Table 1: List of measures included on 5Essentials surveys

Effective Leaders
- Teacher Influence (T)
- Principal Instructional Leadership (T)
- Program Coherence (T)
- Teacher-Principal Trust (T)
Collaborative Teachers
- Collective Responsibility (T)
- Quality Professional Development (T)
- School Commitment (T)
- Teacher-Teacher Trust (T)
Involved Families
- Parent Influence on Decision-making in schools (T)
- Parent Involvement in School (T)
- Teacher-Parent Trust (T)
Supportive Environment
- Peer Support for Academic Work (S)
- Academic Personalism (S)
- Academic Press (S)
- Safety (S)
- Student-Teacher Trust (S)
Ambitious Instruction
- Course Clarity (S)
- English Instruction (S)
- Math Instruction (S)
- Quality of Student Discussion (T)

(S) - Student measure; (T) - Teacher measure

Table 2: Items in measure: Peer Support for Academic Work

How many of the students in your class:

1 None 2 A few 3 Some 4 About half 5 Most 6 All

01 Feel it is important to come to school every day.

02 Feel it is important to pay attention in class.

03 Think doing homework is important.

04 Try hard to get good grades.

Table 3: Items in measure: English Instruction

In your ENGLISH/READING/LITERATURE class this year, how often do you do the following:

1 Never 2 Once or twice a semester 3 Once or twice a month 4 Once or twice a week 5 Almost every day

- 01 Debate the meaning of a reading.
- 02 Discuss connections between a reading and real life people or situations.
- 03 Discuss how culture, time, or place affects an author's writing.
- 04 Improve a piece of writing as a class or with partners.
- 05 Rewrite a paper or essay in response to comments.

Table 4: Items in measure: Math Instruction

In your MATH class this year, how often do you do the following:

1 Never 2 Once or twice a semester 3 Once or twice a month 4 Once or twice a week 5 Almost every day

- 01 Apply math to situations in life outside of school.
- 02 Discuss possible solutions to problems with other students.
- 03 Explain how you solved a problem to the class.
- 04 Write a few sentences to explain how you solved a math problem.
- 05 Write a math problem for other students to solve.

Table 5: Items in measure: Academic Personalism

How much do you agree with the following statements about your class?

1 Strongly disagree 2 Disagree 3 Agree 4 Strongly agree

The teacher for this class:

- 01 Helps me catch up if I am behind.
- 02 Is willing to give extra help on schoolwork if I need it.
- 03 Notices if I have trouble learning something.
- 04 Gives me specific suggestions about how I can improve my work in this class.
- 05 Explains things in a different way if I don't understand something in class.

Table 6: Items in measure: Academic Press

How much do you agree with the following statements about your class:

1 Strongly disagree 2 Disagree 3 Agree 4 Strongly agree

- 01 This class really makes me think.
- 02 I'm really learning a lot in this class.

In my class, my teacher:

1 Strongly disagree 2 Disagree 3 Agree 4 Strongly agree

- 01 Expects everyone to work hard.
- 02 Expects me to do my best all the time.

03 Wants us to become better thinkers, not just memorize things.

In your class, how often:

1 Never 2 Once in a while 3 Most of the time 4 All of the time

01 Are you challenged?

02 Do you have to work hard to do well?

03 Does the teacher ask difficult questions on tests?

04 Does the teacher ask difficult questions in class?

Table 7: Items in measure: Safety

How safe do you feel:

1 Not safe 2 Somewhat safe 3 Mostly safe 4 Very safe

01 In the hallways of the school.

02 In the bathrooms of the school.

03 Outside around the school.

04 Traveling between home and school.

05 In your classes.

Table 8: Items in measure: Student-teacher Trust

How much do you agree with the following:

1 Strongly disagree 2 Disagree 3 Agree 4 Strongly agree

01 When my teachers tell me not to do something, I know they have a good reason.

02 I feel safe and comfortable with my teachers at this school.

03 My teachers always keep their promises.

04 My teachers will always listen to students' ideas.

05 My teachers treat me with respect.

Table 9: Student lexiles by grade

Grade	N	Mean	Std Dev
4	28275	806.61	202.07
5	27492	877.20	193.60
6	27885	919.18	159.53
7	27770	992.81	181.90
8	27127	1023.97	159.2

Table 10: Survey measure text lexile

Measure	Lexile	Typical Grade Level
Student-Teacher Trust	540	2nd
Safety	570	2nd
Academic Press	640	3rd-4th
English Instruction	880	6th-7th
Math Instruction	910	6th-8th
Academic Personalism	920	6th-10th
Peer Support for Academic Work	940	7th-10th

Figure 1: Individual-level reliabilities

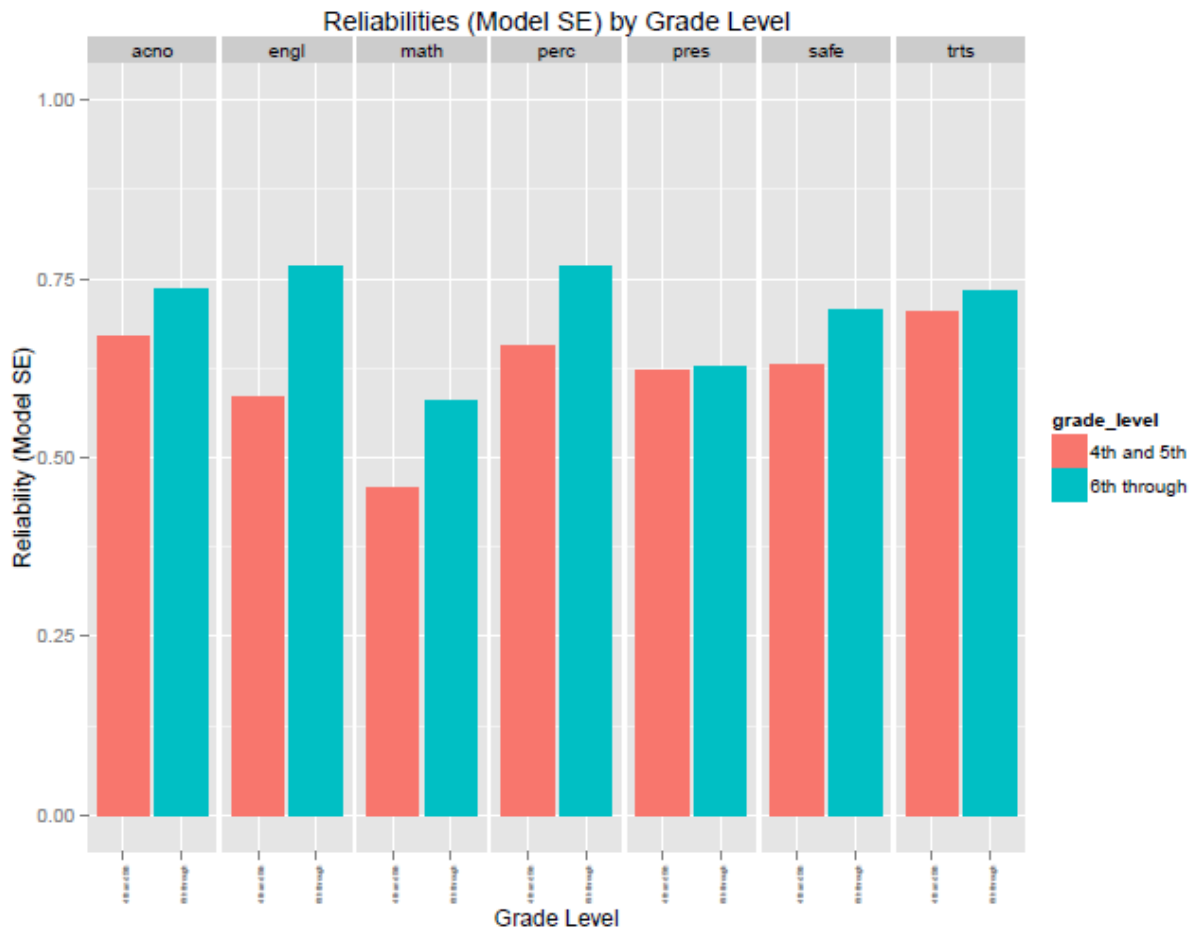


Figure 2: School-level reliabilities

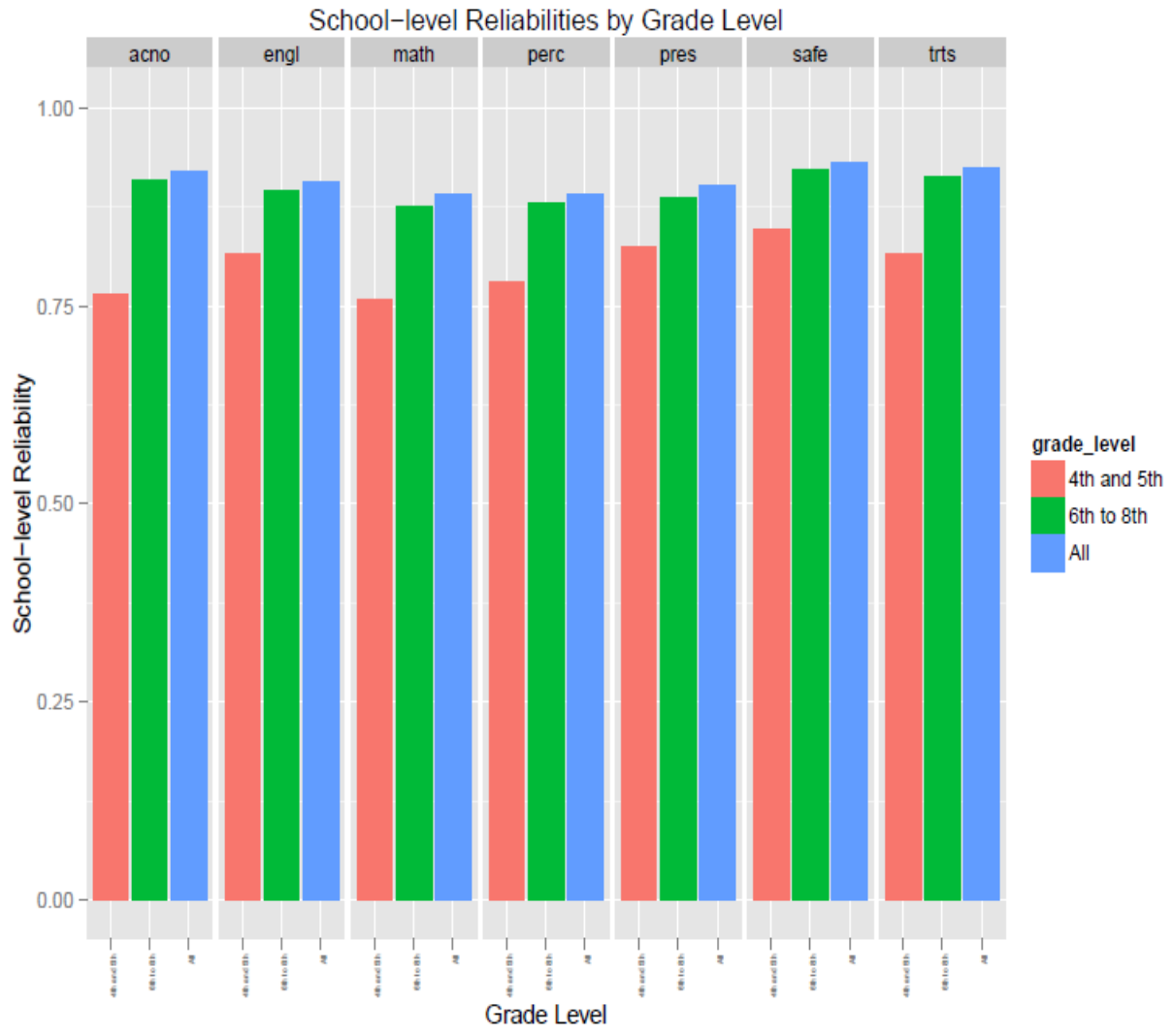


Figure 3: Intra-class correlation

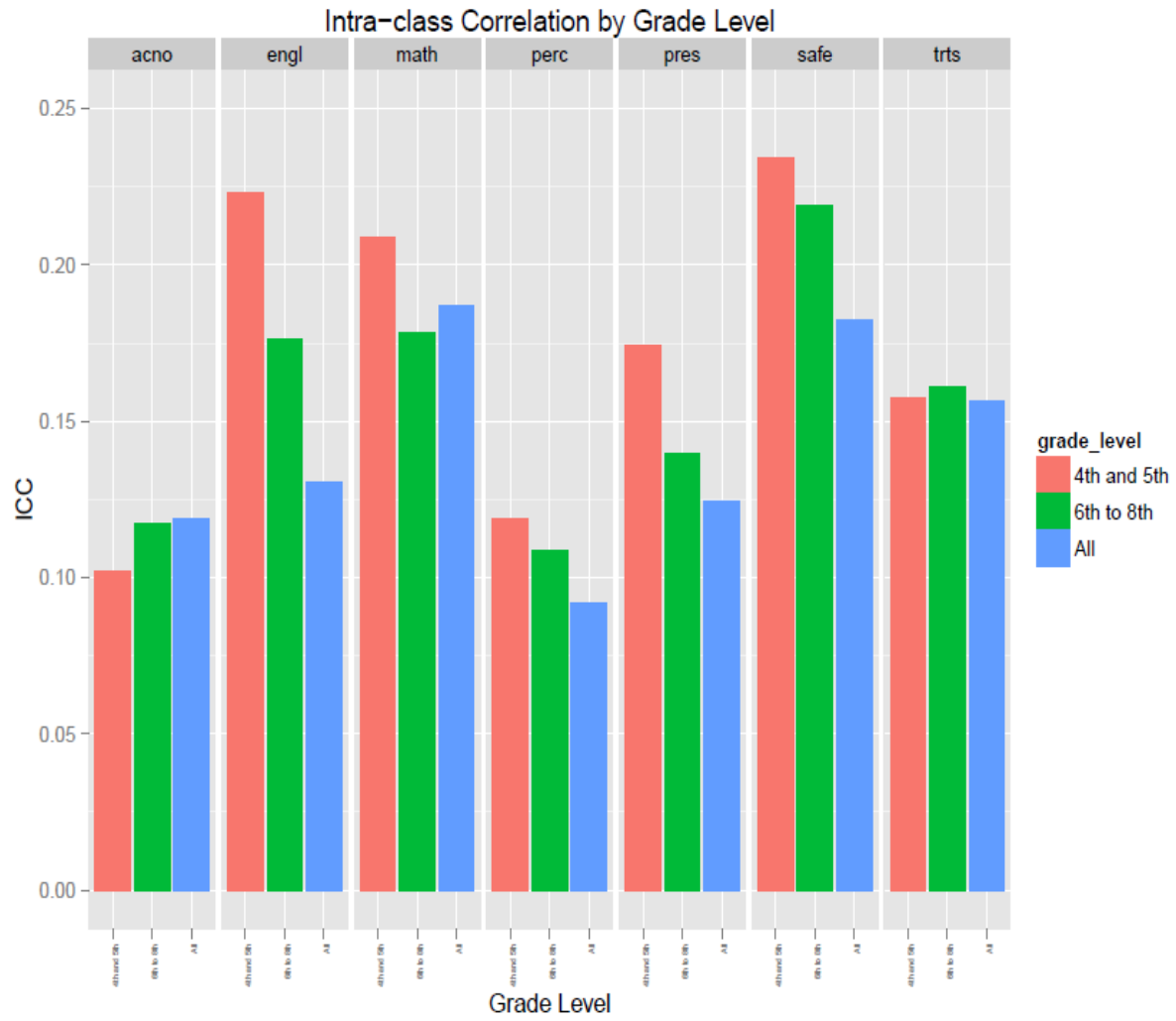


Figure 4: Fit categories by grade for Peer Support for Academic Work

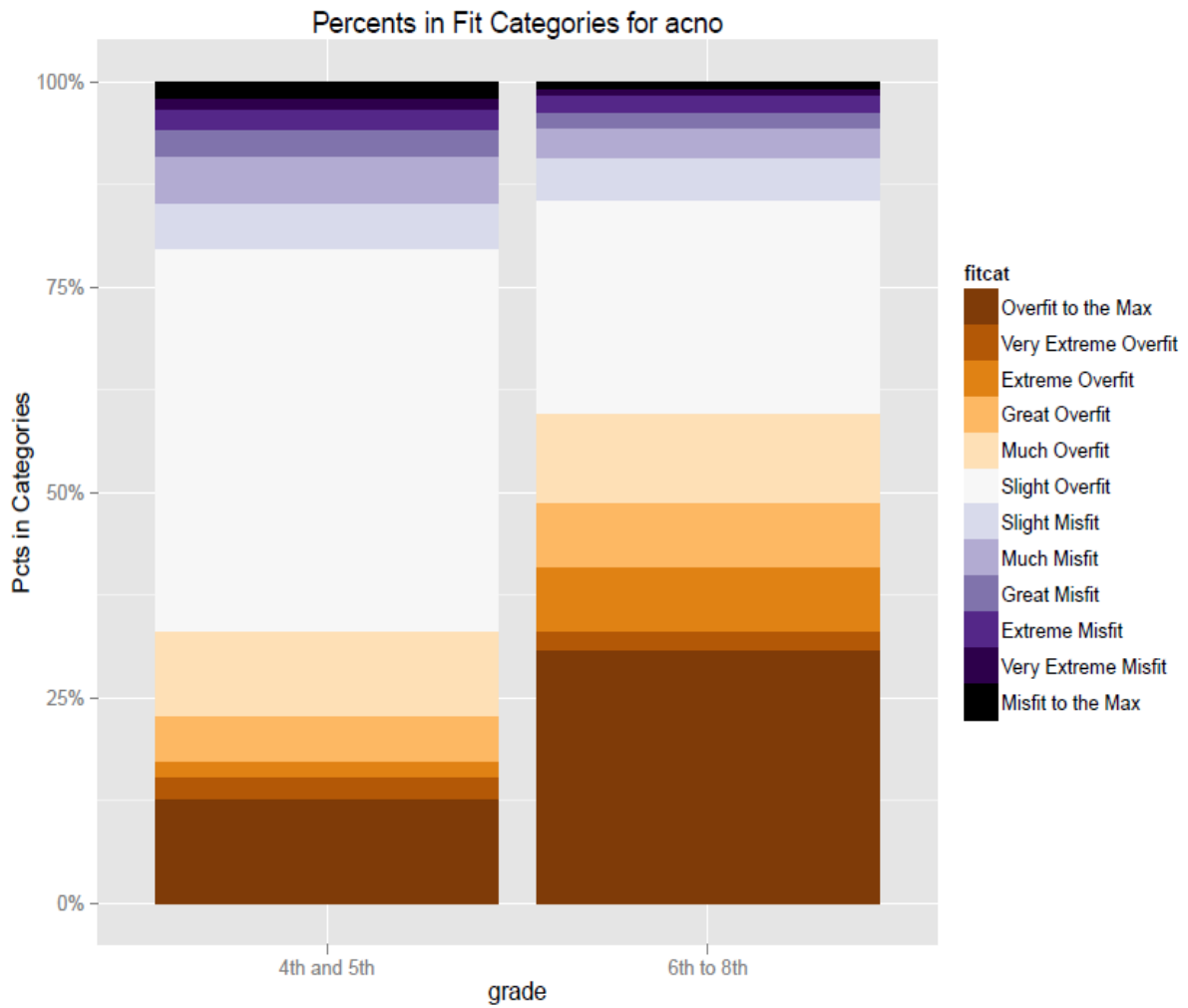


Figure 5: Fit Categories by Grade for English Instruction

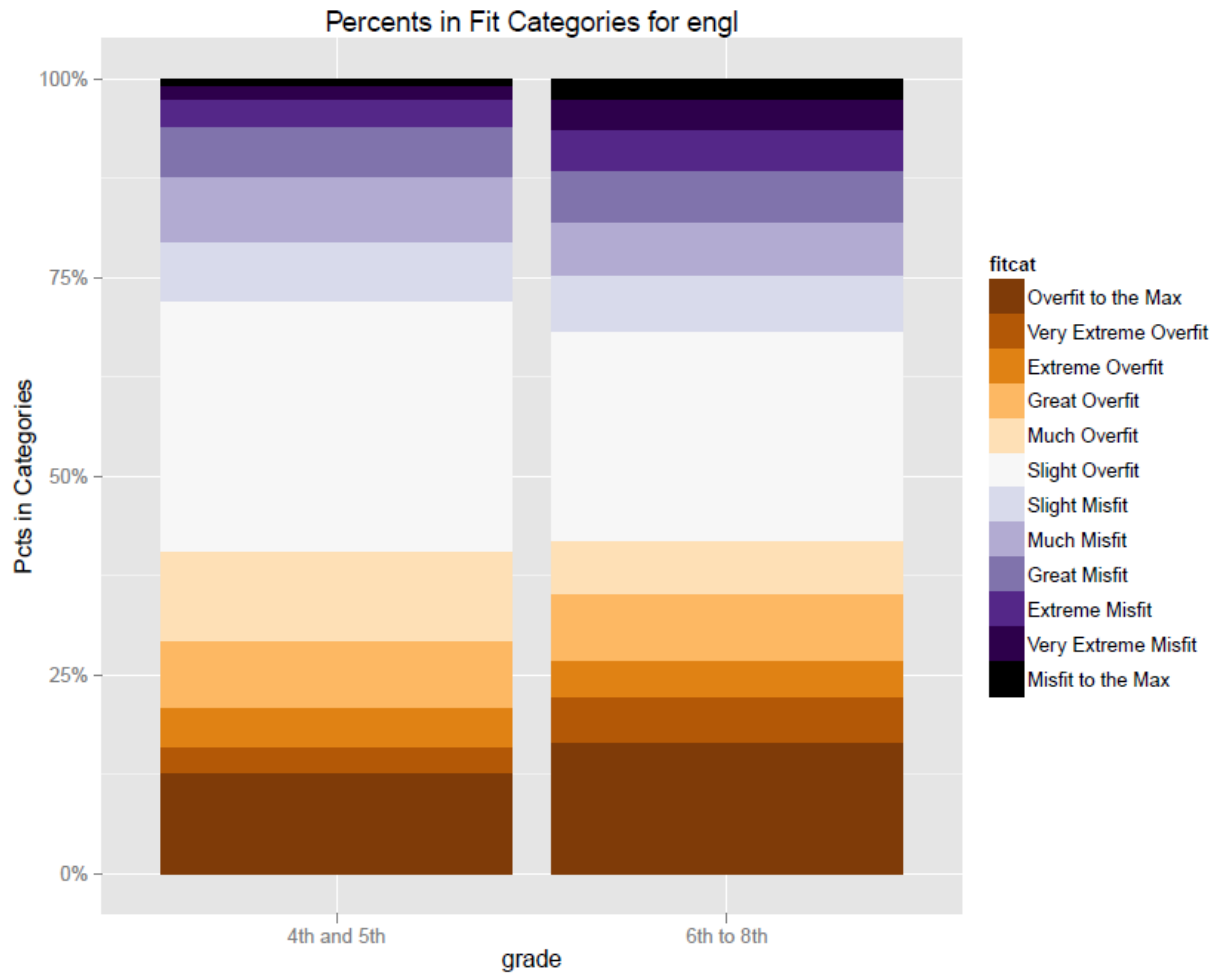


Figure 6: Fit Categories by Grade for Math Instruction

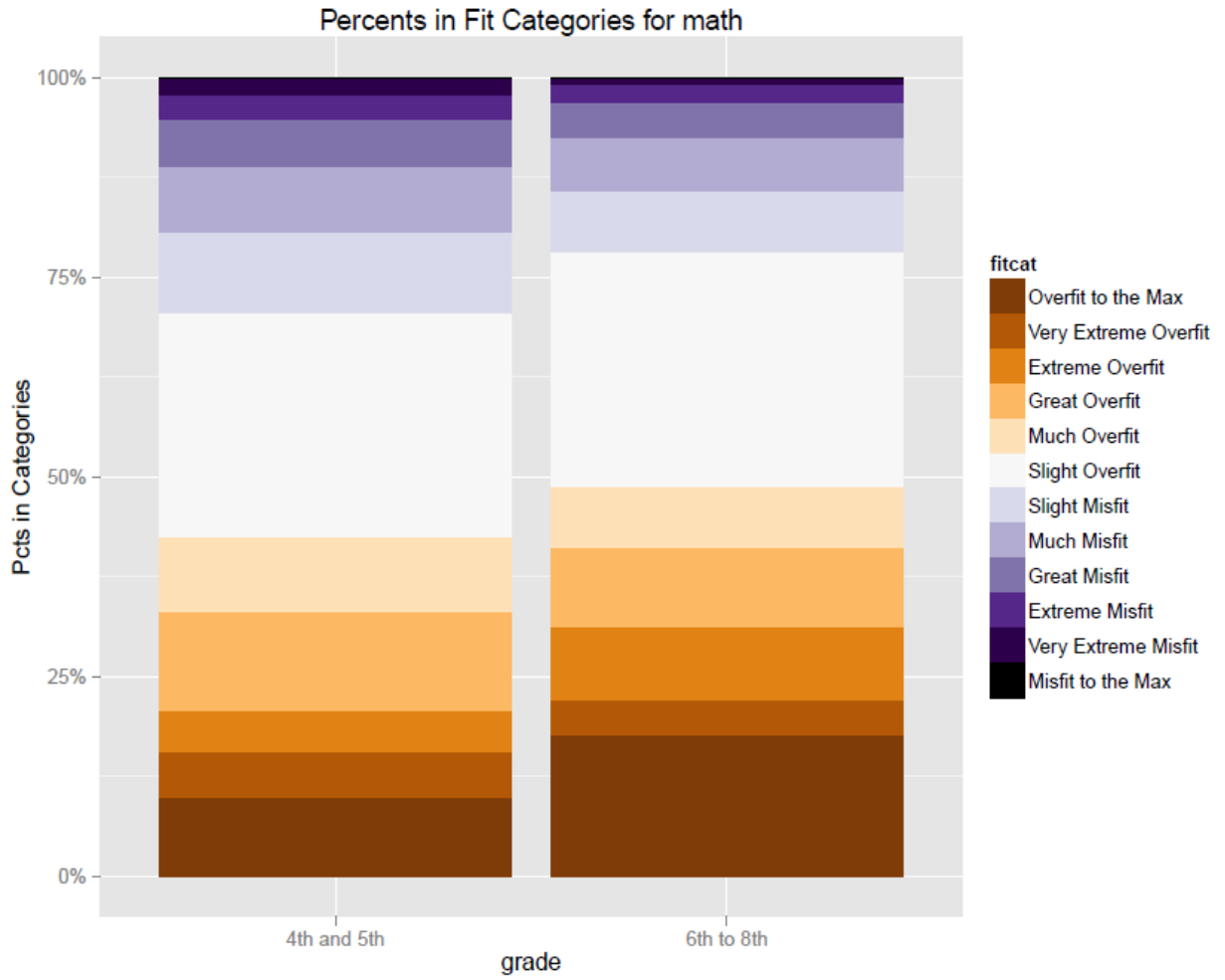


Figure 7: Fit Categories by Grade for Academic Personalism

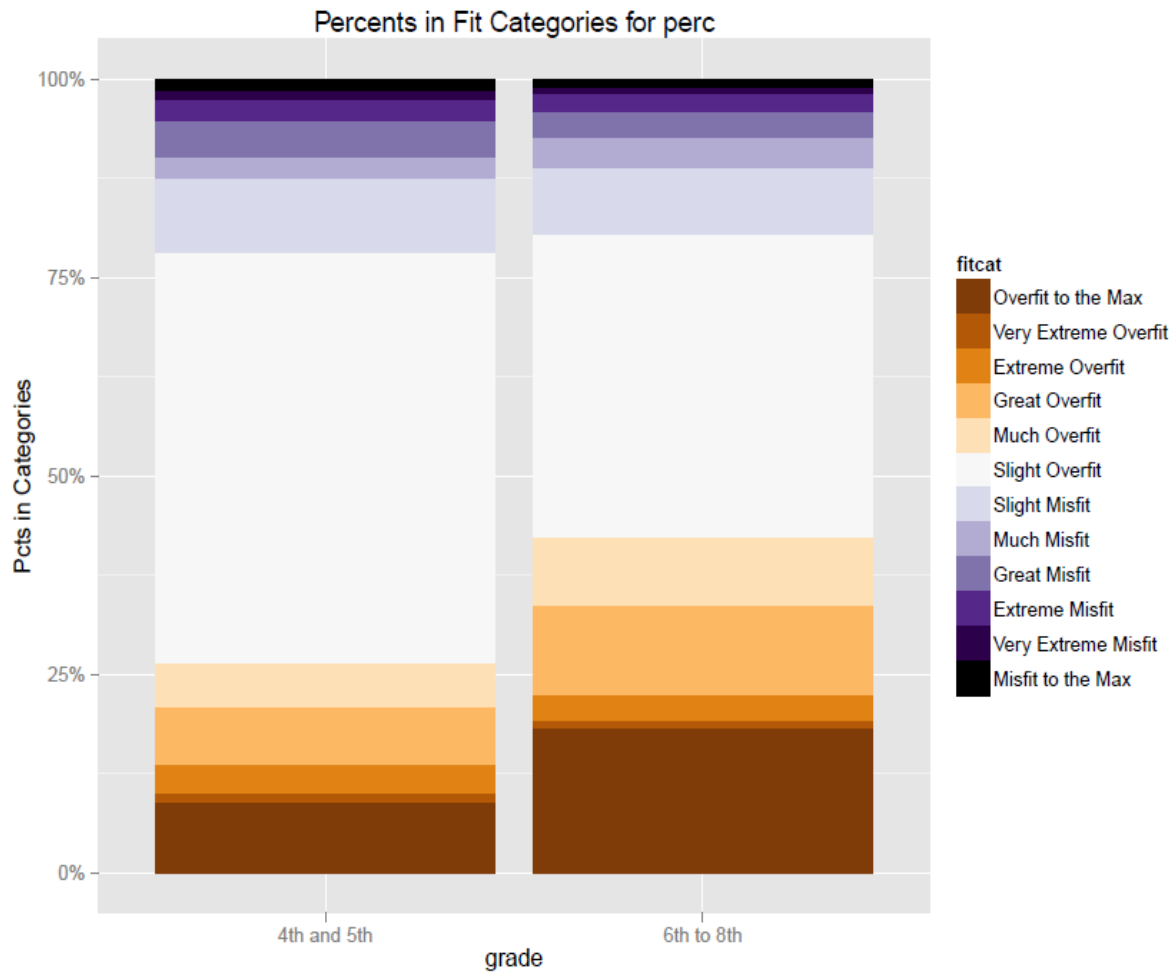


Figure 8: Fit Categories by Grade for Academic Press

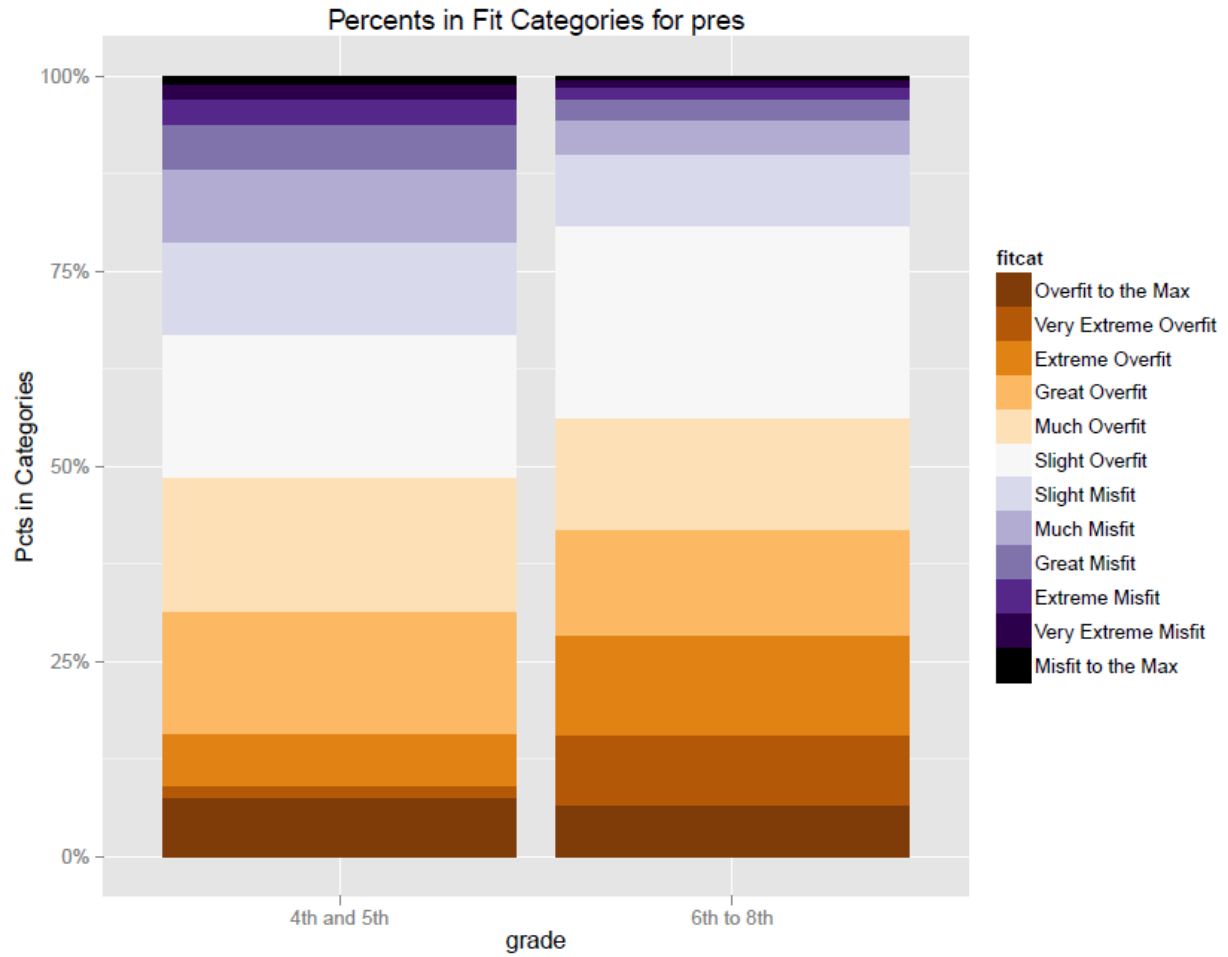


Figure 9: Fit Categories by Grade for Safety

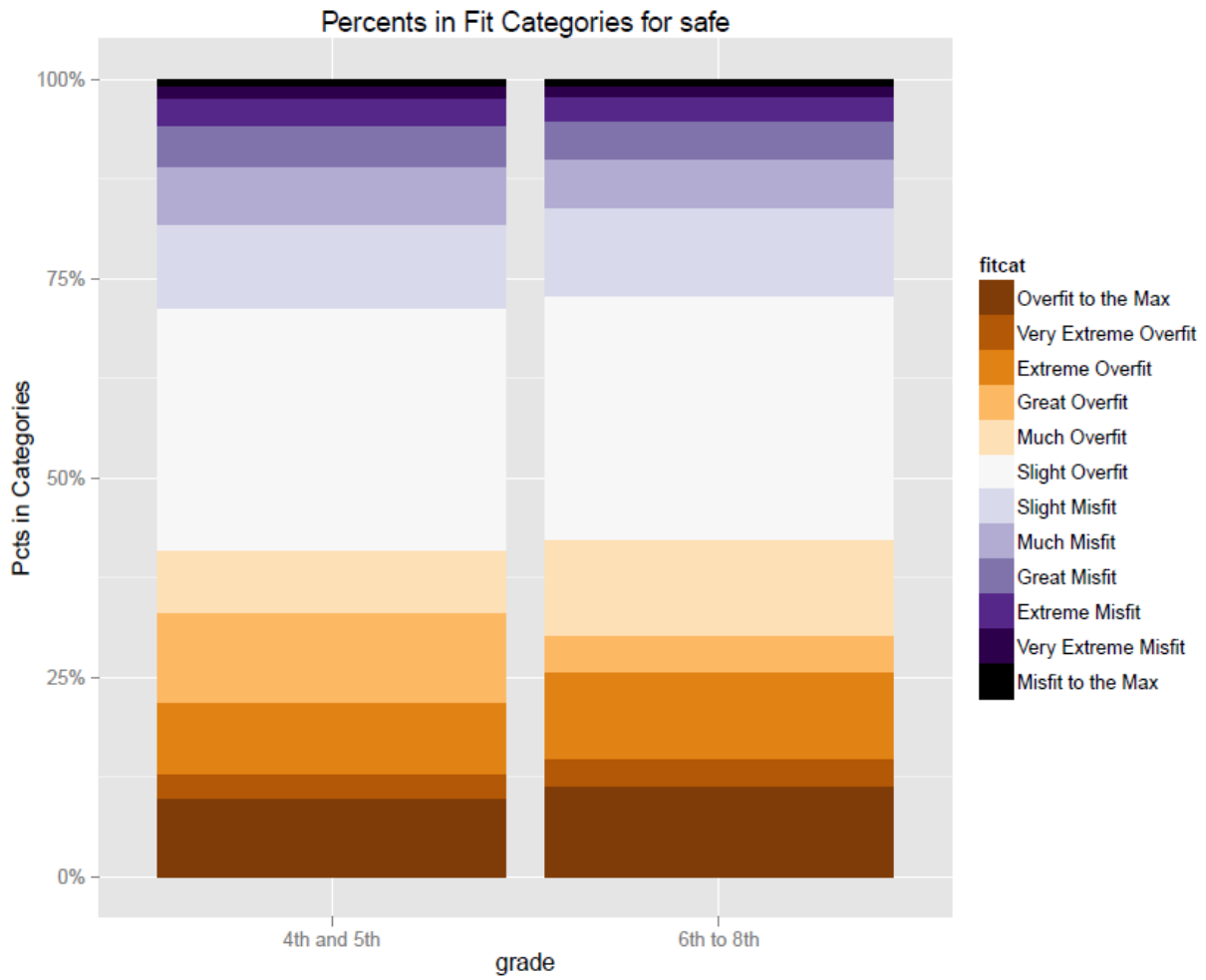


Figure 10: Fit Categories by Grade for Student-Teacher Trust

