

Toward Better Report Cards

Grant Wiggins

The report card should, above all else, be user-friendly: Parents must be able to easily understand the information it contains.

I have heard the following story many times during the past decade: A faculty redesigns its report card, trying to provide more helpful information to students and parents. While time-consuming to compile, the narrative-based system has a great advantage: It plays down naked scores and crude comparisons of students. Nevertheless, at the first conference, almost every parent asks "OK, but how is my child doing?"

This little tale has two vital implications for faculties knee-deep in reform: The customers (parents) are *always* right, regardless of the report writers' intentions; and they typically require more comparative and background information than teachers prefer to provide.

To know how a child is doing, the parents need a context: compared to *what*? No matter how detailed, a narrative can never tell us whether language that describes, praises, and criticizes is relative to our expectations for the child, classroom norms, or absolute high standards of achievement. Adding a *single* letter grade helps very little: the parent still does not know whether the grade represents relative or absolute achievement. Some schools do give comparative data about individual performance against local norms, and many letter grades implicitly provide such a comparison. Yet, mere norms mislead: maybe the class is so heterogeneous that we compare apples and oranges without shedding light on whether each student's performance level is appropriate; or maybe the student's performance *is* good compared to students in that class, but mediocre compared to students in the best classrooms in the region.

The problem with our report cards is that grades and comments are always encoded and not standard-referenced. Current report cards say too little about the *specific* tasks the student has actually done or not done, and to what specific and *verifiable* level of performance. And they say too little about progress toward exit-level standards. What has the child actually accomplished or not accomplished? Is the child on course to perform well at the next school and meet district, regional, and national standards? We need to provide more contextualized, credible, verifiable, and—above all—honest information in report cards.

New Approaches

I propose six new approaches:

1. A clear distinction between standard-referenced and norm-referenced achievement in reports. How is Johnny doing—not just against local norms, but also against credible regional or national standards?
2. A system that sums up the data in two kinds of teacher judgments: judgments about what I term progress (toward uniform K–12 exit standards in each area), and growth (against our expectations for each student).
3. A longitudinal reporting system that charts achievement against exit-level standards, so that a 3rd grader knows how he or she is doing against 5th grade and (sometimes) 12th grade standards, just as we find in performance areas like chess and diving.
4. Many more "sub-grades" of performance. The report should identify strengths and weaknesses in the diverse priority areas, topics, skills, and understandings that make up a subject.
5. Accurate distinctions between the quality of students' work and the sophistication (or degree of difficulty) of their work.
6. An evaluation of the student's intellectual character—habits of mind and work—based on performance and products. The report highlights teacher judgments about the dispositions that are essential to successful higher-level work and routinely found on college reference forms and personnel records (for example, persistence, attention to detail, and open-mindedness).

Such information will make reports more valid, but not necessarily more informative to the parent, who might ask: "What specifically did my child do to earn the grade? What does he or she have to do to earn a higher grade?"

Thus, I also call for a set of background materials that includes such artifacts as anchor papers, performance samples, rubrics, and teacher commentaries so that students and parents can verify the report, not just note it.

I am not advocating the end of the use of letter grades on report cards. Letter grades per se are not the problem. Using a *single* grade with no clear and stable meaning to summarize all aspects of performance is a problem. We need more, not fewer grades; and more different *kinds* of grades and comments if the parent is to be informed.

All the ideas stem from two overarching values: A school's reporting categories and feedback are only as good as the assessment system from which they are drawn, and honesty is the best policy in reporting.

Ease of Translation

A report card summarizes student performance. Grades or numbers, like all symbols, are an efficient way to do this. Because the parent cannot be expected to review all the student's work and arrive at all appropriate meanings, the professional's job is to make meaning from the work and present facts, judgments, and prescriptions in a user-friendly form. The key to report card change, then, is to ensure that grades, scores, or any other system can be effectively translated by parents. Comments may well be desirable. They provide rich, insightful detail, but they do not replace the facts about performance that are summarized in scores and grades.

The problems of report card vagueness and unreliability are not inherent defects of our letter grade system. Some symbols have deep and obvious meaning, such as the best company logos or a filled-in baseball scorecard; others do not. Grades are clear if clear standards and criteria are used, in a consistent way, by each teacher. Grades are unclear if they represent idiosyncratic values and inconsistency from teacher to teacher. Narrative comments don't change this fact. Who can be sure what the teacher means by "Johnny has made great progress and is a delight to have in class"?

But pity our teachers. For most, the giving of a grade is always an ugly compromise. It must summarize their view of student performance measured against their expectations, yet somehow relate to classroom and perhaps regional standards. And, it typically involves factoring in judgment about effort and attitude. In other words, grades rarely represent what the parent thinks: achievement. Our grading system is confusing to report readers who believe that grades are earned in reference to fixed standards, not individualized expectations.

Scores vs. Grades

Our first challenge, then, is to help parents know how their child did from two perspectives: How did the child do when compared against authentic and valid standards? And, how did the child do when we consider all the unique factors that lead to teachers judging performance in light of reasonable expectations? I want to reserve "scores" for the former judgment, and "grades" for the latter.

How do "scores" differ from "grades"? Think of scores as pure performance data, such as those generated on standardized criterion-referenced or performance tests in such areas as diving or music. No mitigating factors are considered; the performance is scored in reference to fixed criteria and standards, through rubrics, exemplars, anchors, or specifications (for example, 100 words per minute in typing). A standard is a standard; in standard-referenced scoring, there should be no predictable "curve." Any performance test (not just a multiple-choice test) should yield valid and reliable scores, common criteria, and standards based on consistent administration. The letter grade should be a separate judgment, designed to reflect reasonable expectations for each student.

Norm-referenced scores are worth reporting. They, too, are data without personalized judgment. It is a fact that you are in the top quartile for your class; it is a fact that your score of 6 on a 6-point scale for writing portfolios was earned by only 4 percent of your class. Such a combined standard-referenced and norm-referenced system was used in Toronto as part of its new Benchmarks program. The rubrics for a Level Five (highest) and a Level Two score on an 8th grade oral performance task read as follows:

Level Five : The student is aware of the importance of both content and delivery in giving a talk. The content is powerfully focused and informative. The issue is clearly defined, and detail is judiciously selected to support the issue. The talk is delivered in a style that interests and persuades the audience. Questions, eye contact, facial expressions, and gesture engage the audience. The student displays evidence of social, moral, and political responsibility, and offers creative solutions. Causes and effects are elaborated. The second version of the talk reveals significant changes based on revision after viewing. The student may make effective use of cue cards. The student is confident and takes risks. (Achieved by 8 percent of the students.)

Level Two : The student's talk contains some specific information with some attempt at organization. The main idea is unclear, and facts are disjointed. Some paraphrasing of text is evident. The student uses no persuasive devices, has little eye contact or voice inflection, and does not take a clear stand on the issue. The delivery is hesitant and incoherent. Little improvement is shown in the talk after watching the first version. The student demonstrates little confidence. (Achieved by 22 percent of the students.)

A letter grade is given in addition to this information. It represents how the pure score translates into a personalized judgment, based on expectations: "Given your particular experience and abilities, Johnny, the work this term represents a fine performance: B+." Or, "Given your extensive experience and my expectations, formed by many prior students like you, Jinni, your relatively good scores are not as high as they could be: C+."

A letter grade supplements, not replaces, the achievement reports (performance task scores, exit-level test scores, and so on). It provides an evaluation of the achievement as viewed from the perspective of personal growth.

Longitudinal Reporting

Our interest should not only be in valid scores, but in feedback given over time that measures progress toward a performance goal. That's very different from each teacher separately judging a student's personal growth over the course of each year, as is now typically the case.

Progress is an *objective* measure of performance gains made over time on a standard-referenced longitudinal scale. It is measured "backwards" from a desirable destination—the standard. It would be silly to say "I have been driving for hours and have made great progress" if I am only 15 miles from home and still 50 miles from Buffalo. Similarly, in the Toronto assessment, student performance is measured backwards from the standard (Level Five) over a multi-year period *irrespective of prior experience or our current expectations of the student*.

Growth, by contrast, represents a judgment about whether current performance falls short of, meets, or exceeds our expectations for that particular student at that time. In a sense we look to the past, not forward to the destination (standard).

Maybe Jimmy is a novice traveler and was far less comfortable in public speaking than we thought he would be. Maybe Julie has become much more comfortable in the journey but has not yet made discernible progress toward exit standards. Maybe Joey has done a bit less than we expected but has made great progress, given his native talents in speaking. A student can do a lot of growing but make little progress, and vice versa. That's why we should report both.

In assessing progress over time, we are objectively assessing the trend: Is a student on course or not to get to the destination in time? In other words, progress judgments are predictions, not value judgments. Using historical data and professional judgment, the teacher makes a prediction about the student's likelihood of meeting exit-level standards. We should also report norm-referenced data: Given the progress and levels of achievement, how do those rates and levels compare with the student's classmates and with classmates from past years?

By reporting progress, we greatly reduce inappropriate norm-referenced comparisons and improve feedback to our less able students: Being "slow" or "behind" is no longer highlighted. Rather, we chart the, perhaps modest, gains over time and worry only whether the trend is a happy one. (For example, will the student likely graduate if the rate of progress continues, given past rates of similar students?)

student likely graduate if the rate of progress continues, given past rates of similar students?)

Such a standard-referenced system, used over the student's career, is both feasible and desirable. We see the value clearly in such systems as karate belts and computer game levels. The task is to devise standards, criteria, and benchmarks to describe and chart student performance. Figure 1 shows how this measurement of progress might look and how normative data about progress can be briefly charted at the bottom. (Note that in this example, the standards refer to elementary exit-level standards. There is as yet no attempt in this district to report progress against 12th grade standards in the lower school units. The "advanced" and "proficient" standards here refer to each school's exit standards.)

Figure 1. Cherry Creek School District Polton Community Elementary School Fairplay Progress Report (Language Arts Section)

Student Name _____ Grade 3 ____ 4 ____ Teacher _____ School Year _____			
Performance-based graduation requirements focus on student mastery of the proficiencies. The curriculum and written progress report are geared toward preparing students for this task. A date (for example, 11/93) indicates where a student is performing on a continuum of progress based on the fifth-grade exit standards.			
	Basic	Proficient	Advanced
Language Arts Proficiency 1. Listens, interpreting verbal and nonverbal cues to construct meaning.	Actively listens, demonstrates understanding, and clarifies with questions and paraphrasing.	Actively listens for purpose, demonstrates understanding, and clarifies with questions and paraphrasing.	Actively listens for purpose, demonstrates understanding, clarifies with questions and paraphrasing, classifies, analyzes, and applies information.
Language Arts Proficiency 2. Conveys meaning clearly and coherently through speech in both formal and informal situations.	Appropriately speaks to inform, explain, demonstrates, or persuade. Organizes a speech and uses vocabulary to convey a message.	Appropriately speaks to inform, explain, demonstrates, or persuade. Organizes a formal speech and uses vocabulary to convey a message.	Appropriately speaks to inform, explain, demonstrates, or persuade. Organizes a formal speech with details and transitions adapting subject and vocabulary. Uses eye contact, gestures, and suitable expression for an audience and topic.
Language Arts Proficiency 3. Reads to construct meaning by interacting with the text, by recognizing the different requirements of a variety of printed materials, and by using appropriate strategies to increase	Reads varied material, comprehends at a literal level. Recalls and builds knowledge through related information. Begins to use strategies to develop fluency, adjusting rate when reading	Reads varied material, comprehends and draws inferences, recalls and builds knowledge through related information. Applies strategies to increase fluency, adjusting rate when reading different material.	Reads varied material, comprehends literally and interpretively. Synthesizes and explores information, drawing inferences. Critiques author's intent, analyzes material, analyzes material for meaning and value. Applies strategies to increase fluency, adjusting rate when reading different material.

Comprehension.	different material.		
Language Arts Proficiency 4. Produces writing that conveys purpose and meaning, uses effective writing strategies, and incorporates the conventions of written language to communicate clearly.	Appropriately writes on assigned or self-selected topics. Clear main ideas, few details. Weak elements in the beginning, middle, end. Sentence structure lacks variety and contains errors.	Appropriately writes on assigned or self-selected topics. Clear main ideas, interesting details, clear organization, sequencing, varied sentence structure, edits to reduce errors. Appropriate voice and word choice.	Appropriately writes on assigned or self-selected topics. Connects opinions, details, and examples. Effective organization and sequencing, meaningful sequencing structure, edits to eliminate most errors. Appropriate voice and word choice.

Figure 1. Cherry Creek School District Polton Community Elementary School Fairplay Progress Report (Language Arts Section) (continued)

As compared to the class in the area of Language Arts, your child...			
Marking Periods	1	2	3
Displays strong performance			
Demonstrates appropriate development			
Needs practice and support			
<i>Editor's Note:</i> The teacher places a check in one box per marking period to indicate child's status in language arts.			

Similar information is given on math and science performance, as well as normed information on a host of intellectual habits and social conduct. These scales are used repeatedly so that progress can be visually represented by movement on the scale over time. And, content standards are not lost: the report describes the books read, specific performance tasks mastered, and course-specific major assignments completed.

Norms have their place here, as noted in my opening story. The staff members of Polton Community School were first going to report only the achievement level on each continuum. Sure enough, the parents demanded more norm-referenced information. The box at the bottom, where the child's performance is compared to that of the class, was added to respond to parent feedback.

Scales are not self-explanatory, however. The descriptors on the report card for "Basic," "Proficient," and "Advanced" are helpful, but ultimately are abstractions. More content-specific descriptors are needed to provide a sense of just what the child can and cannot do without overwhelming the parent with data.

Consider, for example, the report used in South Brunswick, New Jersey, to chart progress toward sophisticated literacy in reading in the early school years. The following excerpts show two of the six development levels of K–2 children in reading and writing:

1 - EARLY EMERGENT READER: Displays an awareness of some conventions of reading, such as front/back of books, distinctions between print and pictures. Sees the construction of meaning from text as "magical" or exterior to the print. While the child may be interested in the contents of books, there is as yet little apparent

attention to turning written marks into language. Is beginning to notice environmental print.

4 - **ADVANCED BEGINNING READER:** Starts to draw on major cue systems: self-corrects or identifies words through use of letter-sound patterns, sense of story, or syntax. Reading may be laborious, especially new material, requiring considerable effort and some support. Writing and spelling reveal awareness of letter patterns. Conventions of writing such as capitalization and full stops are beginning to appear.

Included in the South Brunswick system is an approach to spelling assessment and reporting that honors the idea of reporting progress “backwards” from a standard. Words are not merely spelled correctly or incorrectly and reported as scores on spelling tests. Teachers in the district have learned to chart student progress during the K–2 grades using what they call a “Word Awareness Writing Activity.”¹ The different levels are based on empirically grounded criteria that catalog levels of sophistication in spelling “hunches” (see fig. 2). Now the parents have a clearer perspective on their youngsters’ progress: the charting of scores over time shows progress toward the standards of correct spelling.

Figure 2. A Scoring Chart for Spelling

Look at the child's spelling list. Were most of the spellings *Precommunicative*, *Semiphonetic*, *Phonetic*, *Transitional*, or *Correct*? This is the child's probable developmental level. You might feel that a child truly falls between two of the categories, but try to put in just one check mark per child.

1. **Precommunicative** spellers are in the “babbling” stage of spelling. Children use letters for writing words but the letters are strung together randomly. The letters in precommunicative spelling do correspond to sounds.
2. **Semiphonetic** spellers know that letters represent sounds. They often abbreviate spellings to represent initial and/or final sounds. Example: E = eagle; A = eighty.
3. **Phonetic** spellers spell words like they sound. The speller perceives and represents all of the phonemes in a word, though spellings may be unconventional. Example: EGL = eagle; ATE = eighty.
4. **Transitional** spellers think about how words appear visually; a visual memory of spelling patterns is apparent. Spellings exhibit conventions of English orthography, like vowels in every syllable, correctly spelled inflection endings, and frequent English letter sequences. Example: EGUL = eagle; EIGHTEE = eighty.
5. **Correct** spellers develop over years of word study and writing. Correct spelling can be categorized by instruction levels; for example, correct spelling for a body of words that can be spelled by the average 4th grader would be 4th-grade level correct spelling.

Norms can also be reported: where are the student's peers on the scale? And, we can judge “growth” by giving a letter grade summarizing the teacher's evaluation of progress in light of reasonable expectations for that student.

Think what we would then know about Suzy: she is a 1st grader whose median word awareness assessment score is 2. A 2 seems poor, but we have normative data that show Suzy is ahead of her peers with her score. In fact, we have data and graphs that predict she will be a 5 sooner than most of her peers. We also are impressed by her work habits and the quality of her reading and writing. She has thus earned an A, since she has met all of our expectations and more.²

No Elimination of Grades

.....

The increasing trend to group heterogeneously requires that we clearly distinguish progress from growth, and absolute achievement from norm-referenced grading. It is unfair to both the able and the not-so-able student to offer each only a single grade in which standards, norms, and individual expectations are combined in some unclear way. Fairness demands that less skilled students not have their work compared to their more talented peers. But honesty demands that we report how all students are doing against high, uniform standards.

We rob all children of a successful future if we do not provide them with information about their absolute levels of performance. And, we deceive parents when we represent low-performing students who are highly motivated and working hard as performing at high levels, as often happens when we use the single grade to reward student growth and effort.

Again, the use of a *single* grade to represent achievement, progress, and growth leads to the difficulty of grading fairly. Is it fair or insightful to readers, for example, for a highly-talented student and a special needs student in a mainstreamed class to get the same (single) grade, when readers assume the grade refers to common achievements? On the other hand, why should special needs students be held to the same standard and level of expectation simply because they inhabit the same room?

Yet another problem is that school systems rarely require teachers to agree on the criteria by which similar work products will be judged. The use of the standard curve for giving grades only exacerbates this already bad situation. Now, the single grade is an artifact that bears no obvious relation to performance,

criteria, or standards. That the use of the curve in the classroom is statistically unwarranted doesn't stop many teachers from using some form of it. It allows them to bypass the harder but more fruitful work of setting and teaching performance criteria from which better feedback, learning, and performance would follow. And, it precludes a concern with fairness that would lead us to factor in expectations in a separate grade.

The wrong conclusion from all of this is to throw away letter grades. Sixty years ago, Harvard President Lawrence Lowell argued the case for working for good grades—but where the grades stand for something of clear value and validity:

To chide a tennis player for training himself with a view to winning the match, instead of acquiring skill in the game, would be absurd.... If marks are not an adequate measure of what the course is intended to impart, then the examination is defective (1926).

What is needed is a reporting system that yields a more accurate and rich profile of the student's accomplishments. We have a useful model that summarizes performance data efficiently and offers a brief narrative judgment about the meaning of the data: the baseball card. For each baseball player, we see a brief description of the previous year's performance in data highlighting the many subdimensions of performance: hits, runs, home runs, runs batted in, walks, strikeouts. Subjectivity and judgments about potential or expectations are minimized: These are the raw scores, without explicit meaning.

We can derive much meaning from the numbers, though. Did the ball players play 140 or more games (hence, they were starters)? Were their averages high, compared to other players? We also see the longitudinal trends, since the data are reported for all past years.

Note, though, that evaluation is only implicit in the data; that is dependent upon knowledgeable fans and is subject to surprising disagreement. Only a fan knows that doing well at the plate 3 out of 10 times is a good performance. But even fans disagree as to whether it is "better" to be a .330 hitter who never hits home runs or a .250 hitter who hits 40 home runs.

This example clarifies why the parent *needs* those normative comparisons and teacher judgments cast in letter grades, despite all the data, to place the child's performance in context. It also makes clear why a single letter grade is so unhelpful. Who would feel confident giving a single grade to each ballplayer, given 12 data categories? Such a reduction to a single grade is arbitrary—even if computed "objectively"—whether in baseball or school.

Why would it be arbitrary in baseball? Because runs, hits, and strikeouts are *independent* of one another with no clear or agreed-upon "weight" relative to other data. Some hitters strike out often, but they also hit many home runs and drive in many runs; others hit only singles, but score lots of runs since they are

frequently on base. Some pitchers win many games but have a high earned run average (runs allowed per nine innings); others have the opposite numbers. There is no simple or valid formula for combining all the data; “averaging” all scores to compute a grade “objectively” hides the fact that we have arbitrarily judged each category of performance to be equally valuable.

Why, then, do we arbitrarily average grades and scores in school—where the dimensions of performance are even more complex and diverse—to arrive at a single grade per subject? Problem solving is not research, is not writing, is not discussing, is not accuracy, is not thoroughness, is not mastery of the facts. And, why do we compute averages over the course of a year? One is either achieving at a certain level, or one is not. Why would we use your earlier grades, for example, if you are now performing at a higher level?³

We might begin report card reform, therefore, by building performance subcategories, standards, and benchmarks from the national reports in each subject.⁴ In the National Council of Teachers of Mathematics Standards, for example, performance is divided into mathematical power, problem solving, and so on. These are above and beyond content standards.

Why not encourage all math teachers to disaggregate their letter grades into these separate grades—based on data from tests where rubrics are used for each standard? Why not make sure that English teachers report each student’s performance on different genres of writing, because performance across genres is not constant, as many writing assessments have shown?

More disaggregated scoring—where achievements and progress are separated, and where performance is separated into its many subscores—would increase the incentives of the report for students as well as the clarity of the report for parents. Particular strengths would more likely be revealed; particular weaknesses would more reliably be identified.

Honesty is always the best policy in reporting. The Outcome-Based Education/Mastery Learning system of reducing grades to *A*, *B*, *Incomplete*, or getting rid of grades altogether confuses helpful evaluation with the disincentive of isolated, single grades. Any reporting system must be complete; it must place the student’s performance on a continuum. Eliminating grades lower than *B* makes as little sense as not reporting batting averages under .300.

The complaint that grades lead to invidious ranking is also misconceived. The aim is to better evaluate performance and to efficiently but effectively communicate results. Ranking is a different urge, one *hampered* through the use of multiple grades and scores, as in the baseball card example. The urge to rank is based on a failure to compare performance to *criteria* and, instead, to compare performers to one another. Our urge to reduce things to one number and our over-use of norm-referenced testing and grading are the culprits, not the letter grades themselves.

Quality vs. Sophistication of Performance

Developmental/longitudinal descriptors are not sufficient. We also need to know the quality of the student’s work products and performance. We should, therefore, be developing a reporting system similar to those used in music and gymnastics, where “degree of difficulty” and “quality” are separated.

In music, a fine example can be found in the New York State School Music Association assessment process. All pieces that might be played by any student—either as a soloist or as a member of a small or large ensemble—are ranked by degree of difficulty on a 6-point scale. Once the student has chosen a level of difficulty, he or she is assessed in terms of the various qualities of performance. One naturally expects scores in difficulty to increase over time; there is no stigma to playing Level 1 pieces as a novice.

By not separating the quality of performance from the degree of difficulty of the task, we again blur the meaning of the results. One writer may be much more sophisticated than his or her peers, but also prone to careless mistakes. Another student who is “slow” may produce excellent work for his or her prior experience.

The *B* in English III is in some sense a better performance than an *A* in English I, but not necessarily in the case of each student and our expectations. Again, we see the value of reporting both level of performance and judgment about the result in terms of appropriate expectations.

Implicit in this argument is the need to think of the report card as a mere cover page or “executive

summary," supported by documentation to justify and amplify the meaning of the grades given. Even a baseball card writer assumes that the reader knows what the categories stand for, and what implicitly counts as an excellent performance historically. Few parents understand what constitutes an exemplary performance, however, because few schools disseminate samples of excellent work as a frame of reference.

It stands to reason, then, that the report can only be fully decoded if we provide parents with the rubrics, sample products, and developmental descriptors that are used in assessing student performance. For parents to find the report helpful and credible, they need the tools to verify and understand student scores: the anchor products and the scoring guidelines.

Here is where the narrative report, complemented by student work samples, is so useful. The narrative highlights accomplishments on key tasks, projects, and assignments, and provides a holistic portrait. Parents are able to consider the teacher's judgment in light of the anchor papers and the samples in the student's portfolio. Descriptions of successes and struggles become more meaningful with the backdrop of scores, grades, norms, and work samples.

The narrative, combined with a rubric, is also a good place to make a summary judgment about the student's habits of mind and production. And, if "the customer is always right," we might want to draw from college reference forms in constructing school reports. The rating sheet shown in Figure 3 is part of the private college universal admissions packet used by more than 100 colleges nationwide.

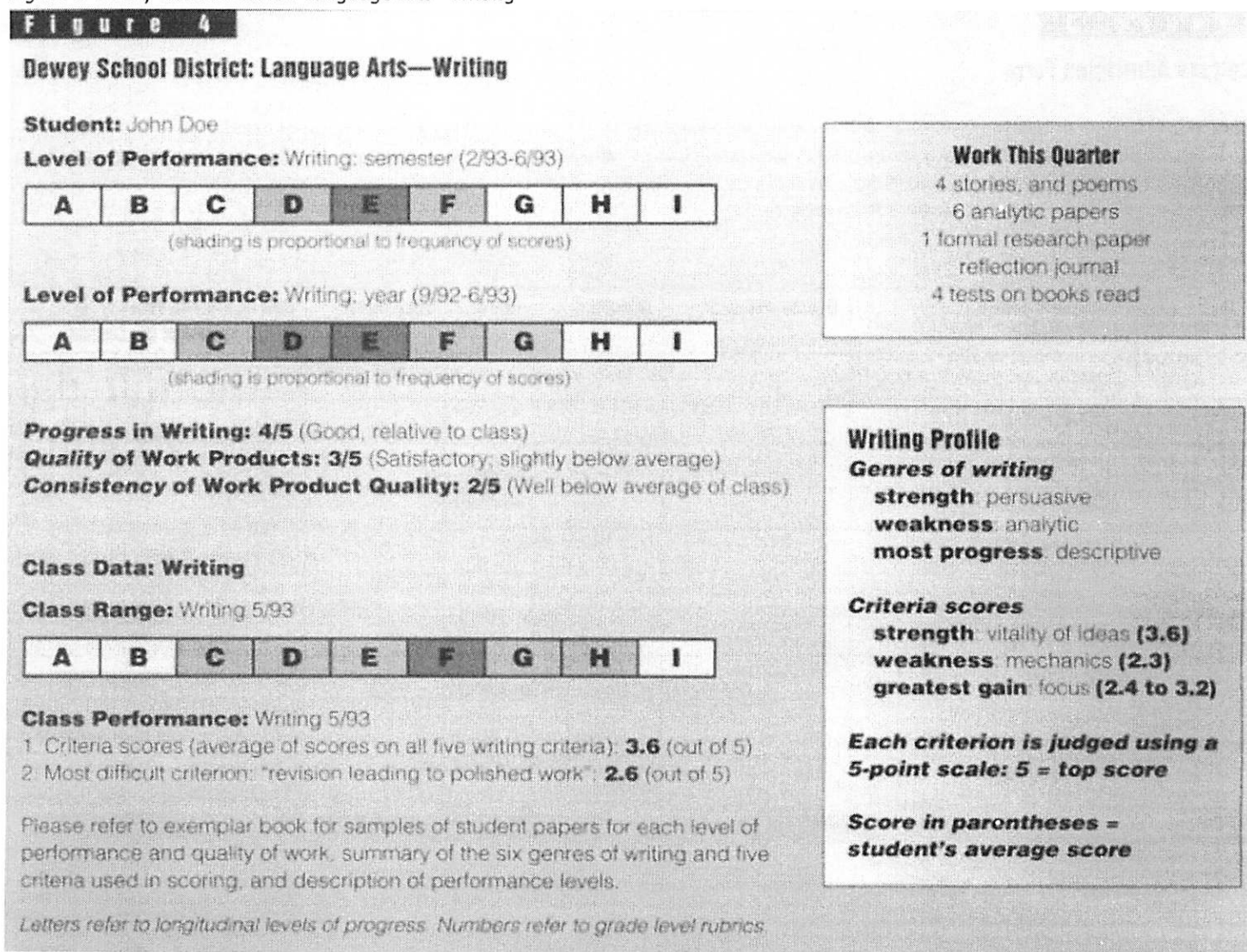
Figure 3. College Admission Form

Please feel free to write down whatever you think is important about the applicant, including a description of academic and personal characteristics. We are particularly interested in the candidate's intellectual purpose, motivation, relative maturity, integrity, independence, originality, leadership potential, capacity for growth, special talents, and enthusiasm. We welcome information that will help us to differentiate this student from others.						
RATINGS						
No Basis	Academic Skills and Potential	Below Average	Average	Good	Very Good	One of the Top Few Encountered in My Career
	Creative, original thought					
	Motivation					
	Independence, initiative					
	Intellectual ability					
	Academic achievement					
	Written expression of ideas					
	Effective class discussion					

	Disciplined work habits					
	Potential for growth					
	SUMMARY EVALUATION					

We at the Center on Learning, Assessment, and School Structure (CLASS) have worked with a few districts to create a document-backed reporting system that notes performance sophistication and quality, using as our guide the work done over the years by the Carleton School District in Ontario, Canada, and similar models. Carleton publishes "exemplar booklets" that provide students, parents, and teachers with the operational standards of assessment: samples of the range of papers, teacher commentaries on those papers, and the rubrics used to score them. Parents are then in a position to actually *verify* (or challenge) the grades given. The report card, then, actually informs and educates the parents. We have developed a mock-up of one page of such a parent report, on a student's achievement as a writer (see fig. 4).

Figure 4. Dewey School District: Language Arts—Writing



Provocative Questions

My conclusions about assessment and reporting can be summarized in a series of questions and answers:

1. *Who is the audience, and what is the purpose, of a report card ?* The audience is the parents, and the purpose is to tell them how their child is performing in terms of standards and expectations. The report card should, above all else, be user-friendly. No parent should review it and say, "I don't have a clear idea about how my kid is really doing." Teachers should thus not be the final judges of what a report card should contain.
2. *What role should letter grades play in report cards ?* They should be used for what they are best at doing: symbolizing the normed judgments a teacher makes about the degree to which a student has met expectations. Grades should not be confused with performance scores, whereby parents learn the student's level of achievement on a continuum ranging from novice to expert.
3. *Is a report card self-sufficient ?* No, it is not. Grades are symbols for verifiable performance and product evaluations. Thus, for the report card to be maximally informative, it must be backed up with work samples, rubrics, anchor papers, and commentary.
4. *What should we report ?* We should relate disaggregated achievements, progress, intellectual character, and specific successes or weaknesses that are highly illustrative of overall performance. "Disaggregated" means, above all else, not using a single grade, score, or description to characterize performance in an entire subject area. Rather, what is wanted, as the Cherry Creek example suggests, is the breakdown of (inherently) complex performance into its many subelements, similar to what we find on the baseball card.

"Progress" is not the same as "growth." Progress is measured backwards from the goal, and growth is typically defined as change in the student. But a student could change a great deal without making much progress toward the standard. "Intellectual character" implies such things as persistence, rigor, craftsmanship—the very qualities found on every college and job recommendation form in America.

References

- Henderson, E. H. (1990). *Teaching Spelling: 2nd Edition*. Boston: Houghton Mifflin.
- Kendall, J. S., and R. J. Marzano. (1994). *The Systematic Identification and Articulation of Content Standards and Benchmarks - Update*. Aurora, Colo.: Mid-Continent Regional Educational Laboratory.
- Lowell, A. L. (1926). "The Art of Examination," *Atlantic Monthly* 137, 1: 58–66.
- Wright, R. G. (May 1994). "Success For All: The Median is the Key," *Phi Delta Kappan* : 723–725.

Endnotes

- ¹ The South Brunswick faculty and Ted Chittendon from the Educational Testing Source developed this activity (see Henderson 1990).
- ² In South Brunswick they do not give such grades. Note that these kinds of assessments against a standard do not favor or undercut any particular teaching or program philosophy.
- ³ Russell Wright (1994) suggests that grades would be more valid if they were based on the *median* grade as opposed to the *mean*. This is a fine idea, given the inherent inconsistency of both performer and assessor, and the need to justify the judgment that a relatively inconsistent student performer can be said to achieve a specific level of performance. Note, though, that his proposal will only work if there are clear and stable performance standards and criteria against which grades are given.
- ⁴ See Kendall and Marzano (1994) for an excellent cross-referenced compendium of all national reports on standards. Note, that most of the reports focus on content standards, not performance standards.

