Formative, Summative, and Performance Assessments

Lacy Gainey

Emat 7050

Fall 2013

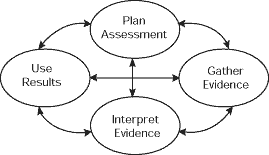
In 1995, the National Council of Teachers of Mathematics (NCTM) published the *Assessment Standards for School Mathematics.* These standards were designed “to complement NCTM’s earlier standards documents, *Curriculum and Evaluation Standards for School Mathematics* (1989) and *Professional Standards for Teaching Mathematics* (1991)” (NCTM, 2013).

“The NCTM Mathematics Assessment Standards include:

1. Assessment should reflect the mathematics that all students need to know and be able to do.
2. Assessment should enhance mathematics learning.
3. Assessment should promote equity.
4. Assessment should be an open process.
5. Assessment should promote valid inferences about mathematics learning.
6. Assessment should be a coherent process.” (NCTM, 2013)

These standards represented a shift in: content, learning, teaching, and evaluation. (NCTM, 2013) Successfully performing rote manipulations is no longer considered to be sufficient. High cogitative demand tasks are now valued. (NCTM 2013, Wilson 1971) Additionally, “high expectations need to be set for all students”. NCTM’s assessment standards can be applied to summative, formative, or performance assessments. “The goal of a summative assessment is to measure the level of success or proficiency that has been obtained at the end of an instructional unit, by comparing it against some standard or benchmark” (Carnegie Melon, 2013). Formative Assessment is referred to as “the gathering and use of information about students’ learning that helps both teachers and students modify teaching and learning activities” (Suurtamm, 2012, p. 29). For performance assessments “students are asked to write their solutions to open problems by using mathematical language and their best representation to communicate how they solved the problem; the decisions they made as they solved the problem; and the connections they made to other problems, mathematics, or other subjects” (Billstein, 1998, p. 284). Performance assessments do vary depending on the criteria being accessed.

NCTM’s *Assessment Standards for School Mathematics* also proposes an assessment process. This process involves four interrelated phases. The phases include: Plan assessment, gather evidence, interpret evidence, and use results. These “phases are interactive and the distinctions between them are blurred” (NCTM, 2013). Figure 1 illustrates NCTM’s (2013) Process Model.

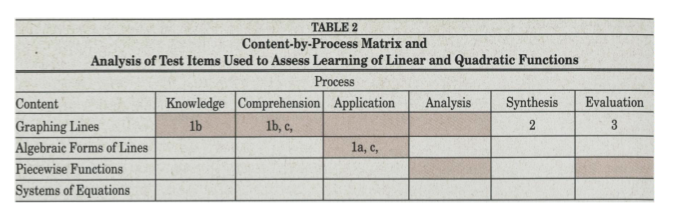
NCTM (2013) notes that its assessment standards and assessment process are “intended to guide teachers and others as they examine current assessment practices and plan new assessment systems”. The assessment process should be expanded on to create assessment systems, frameworks, or models. Wilson (1971) refers to a need for an assessment model by stating, “The goals of mathematics instruction in the secondary schools of this country vary, because there are a multiplicity of outcomes expected from mathematics instruction for any given set of students; because regions, states, schools, or teachers may stress different sets of outcomes; and because mathematics is so universal” (p. 650). Additionally, “It is only reasonable that student should know how they are assessed, what mathematics they will be expected to do, the criteria for judging their performance, and the consequences of the assessment” (NCTM, 2013). While Wilson (1971) and NCTM (2013) referenced a need for an assessment model, they did not specify that there needed to be one model for all types of assessments or even one model for each type of assessment. This paper discusses several different assessment models geared towards summative, formative, and performance assessments

Moskal (2000) created an assessment model by expanding upon the assessment process described in NCTM’s *Assessment Standards for School Mathematics*. Her own teaching experience and the “multiple observations [she] made during the 1994-95 school year of two sixth-grade mathematics classrooms”(Moskal, 2000, p. 192) were the influential factors behind the model. Moskal used the four phases of NCTM’s assessment process to divide the model into four, interconnected sections. Within each section she specified the intended outcome and main actor. Section 1 is titled “Planning”. The intended outcome of this section is an assessment instrument and the main actor is the teacher. The outcome of Section 2: Gathering is student response. The student (main actor) is responding to the assessment instrument. Information is the outcome of Section 3: Interpreting. The teacher interprets the student responses. Lastly, feedback and classroom decision-making are the outcomes of Section 4: Using. The teacher is the main actor in this section. Figure 2 illustrates Moskal’s model. Figure 2 was retrieved from (Moskal, 2000, p. 194).

Figure 2.tiff

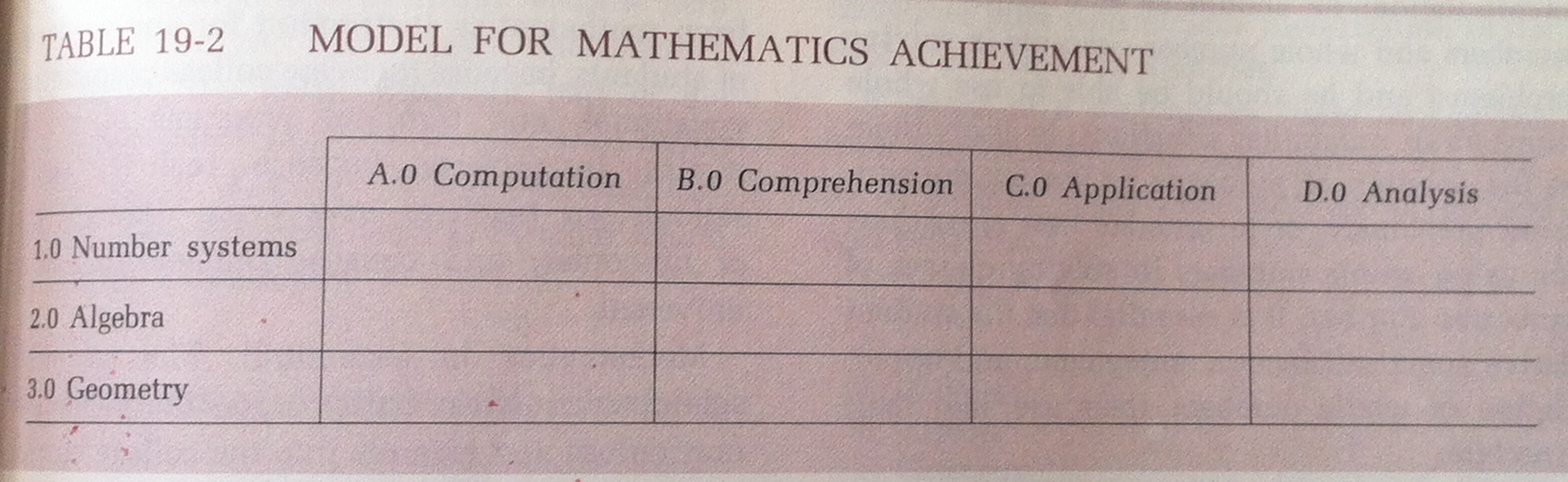
“The arrows suggest that each phase of assessment influences the outcomes in the phases that follow” (p. 193). The cycle described by this model is also continuous. Teachers should use the information they gather from assessments in their future instructional decisions. (Moskal, 2000) This model can be applied to a summative, formative, or performance assessment.

Kastberg (2003) used Bloom’s Taxonomy to create a framework for classroom assessment that is geared towards summative assessments. *Taxonomy of Educational Objectives: Cognitive Domain* (Bloom et al. 1956) contains six levels of behaviors. (Kastberg, 2003) These behaviors include: “Knowledge, comprehension, application, analysis, synthesis, and evaluation” (p. 402). Bloom’s Taxonomy is a cumulative hierarchy. In other words the behaviors are “arranged in order of increasing complexity and [are] cumulative because each behavior [is] presumed to include all of the [previous] behaviors” (p. 402). Kastberg uses a content-by-process matrix to illustrate her framework. The behaviors included in Bloom’s Taxonomy are the processes included in the matrix. The content is dependent on the unit or course that is being taught. Figure 3 illustrates Kastberg’s framework. Figure 3 was retrieved from (Kastberg, 2003, p. 404).

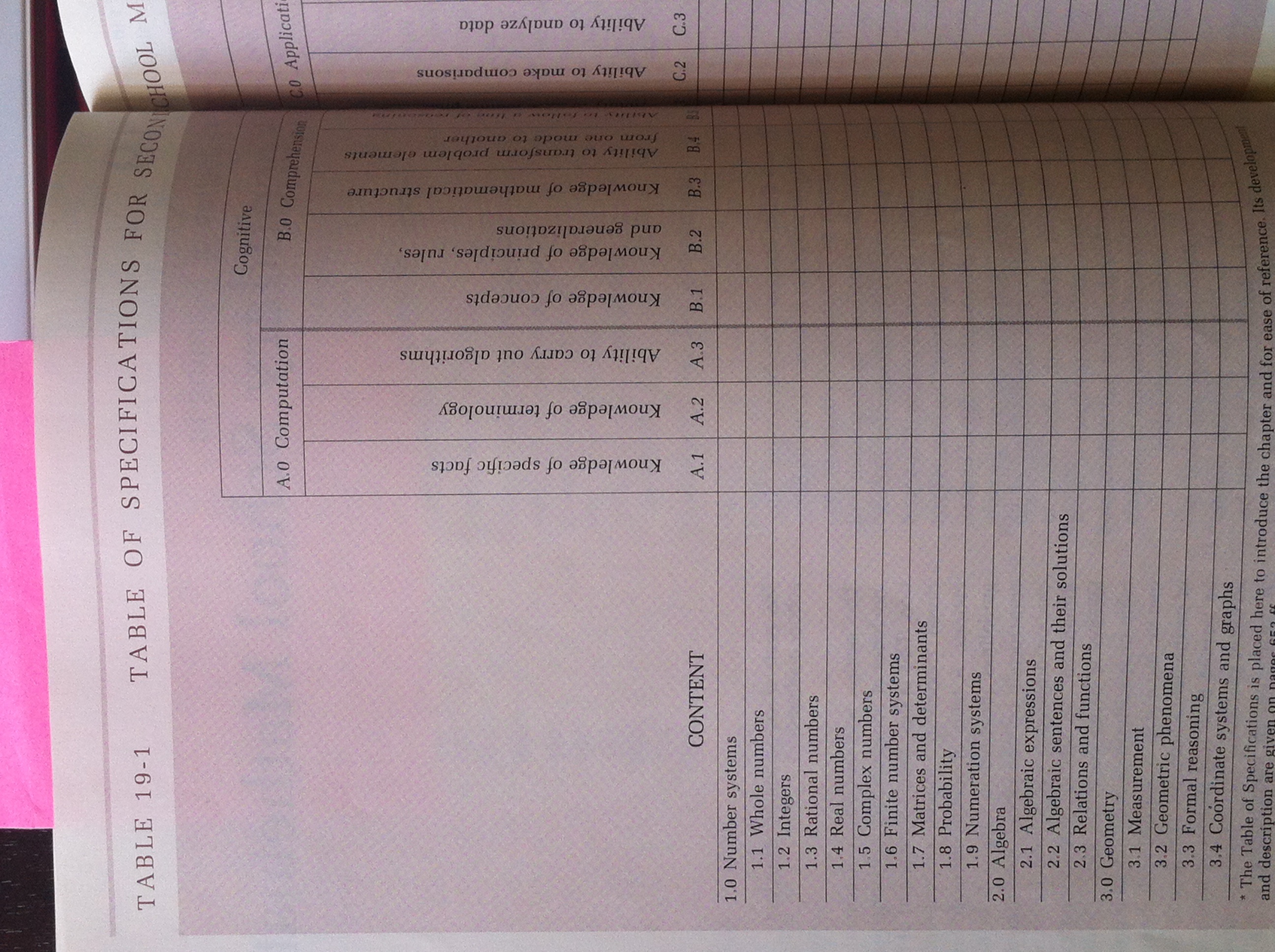


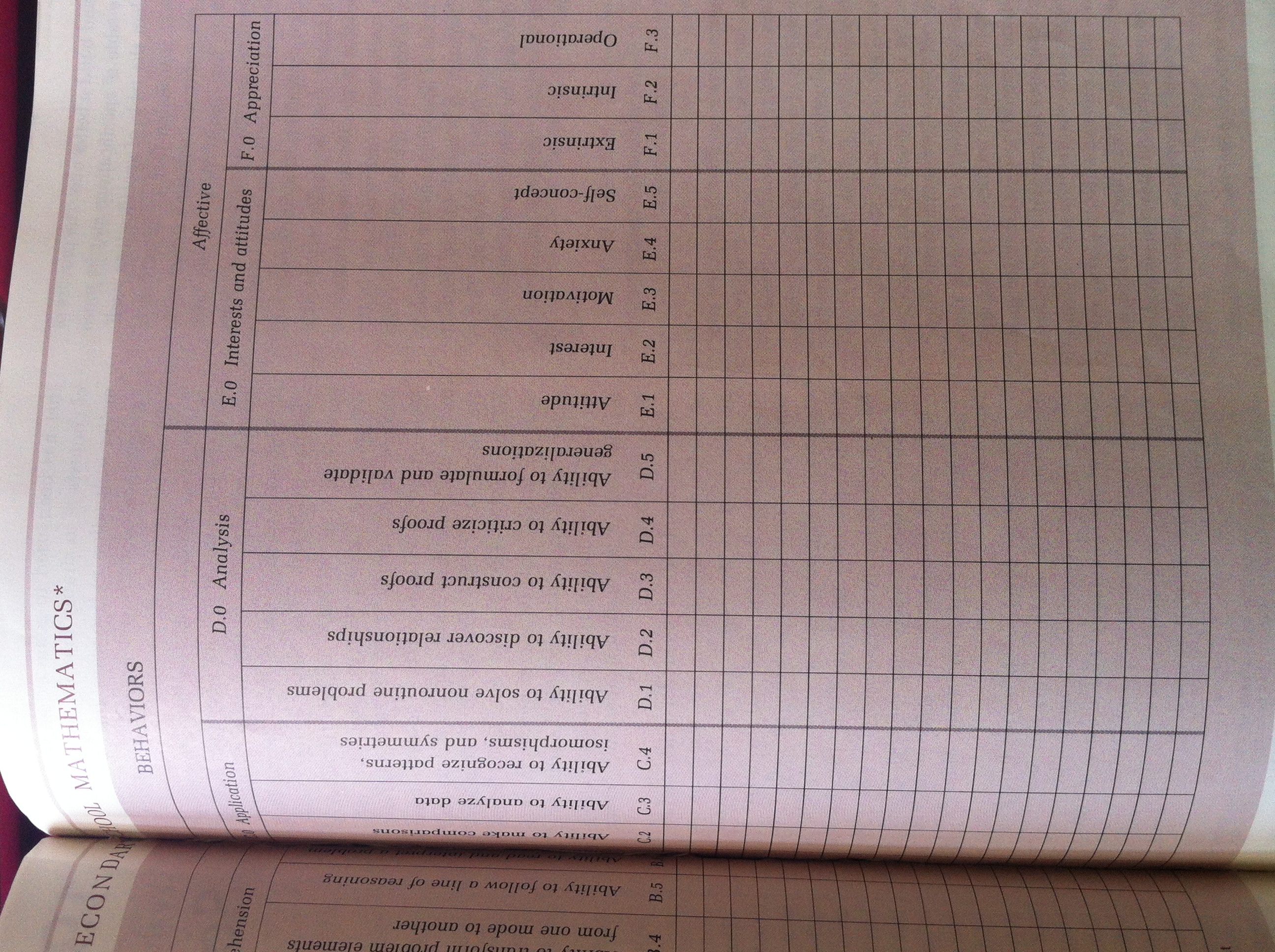
The matrix allows teachers to categorize their assessment items by the content that the item covers and the behavior that the item employs. The matrix gives teachers an overview of their assessment. It allows them to see any gaps in their assessment, in addition to the behaviors they are focusing on. Without this system, Kastberg found herself creating assessments haphazardly. She developed this model out of a need to make her instruction and assessment more systematic, stating “even the best-intentioned evaluation can lead to sparse coverage of what a teacher intends students to learn” (p. 404). In other words, “students have learned that what is assessed and how it is assessed reflect what educators value” (NCTM, 2013). Since assessments convey to students what is valued, it is critical that your instruction and assessments are in alignment. (Kastberg, 2003)

Wilson (1971) created a model that is similar to Kastberg’s framework, but more complex. His model is an expansion of the model developed by the National Longitudinal Study of Mathematical Abilities (NLSMA). (Wilson, 1971) Like Kastberg’s framework, “NLSMA’s model was developed by careful study of the *Taxonomy of Educational Objectives: Cognitive Domain* (Bloom 1956)” (p. 648). Wilson’s final model is a table of specifications for secondary mathematics. This table serves “as an organizational framework to identify outcomes of instruction according to a detailed classification diagramming content by levels of behavior” (p. 693). To develop this table, he first started by creating a model for mathematics achievement. (Wilson, 1971) Figure 4 illustrates Wilson’s model for mathematics achievement. Figure 4 was retrieved from (Wilson, 1971, p. 649).



His model for mathematics achievement has “a content dimension and a behavior, or objectives, dimension” (p. 645). The content dimension includes: Number systems, Algebra, and Geometry. The behavior dimension includes: Computation, comprehension, application, and analysis. Like Bloom’s Taxonomy, these behaviors are “both hierarchical and ordered” (p. 649). Wilson expands upon this model to create a table of specifications for secondary mathematics. He does this by dividing “each of the categories in each dimension into finer subcategories” (p. 649). Additionally, he adds to Affective categories to the behavior dimension. These include interests and attitudes, and appreciation. Figure 5 illustrates Wilson’s Table of Specification for Secondary Mathematics. Figure 5 was retrieved from (Wilson, 1971, p. 646).





Wilson (1971) stresses how this model should be a starting point rather than an ending point. An assessment model not only addresses problems related to evaluation, but problems related to mathematics curriculum and instruction as well. (Wilson, 1971) A teacher should consult this assessment model, or a similar one, during the planning and implementation of their instruction. Referencing an assessment model can help keep teachers on track for reaching their goals involving instruction and assessment. For example, many “teachers often state their goals of instruction to include all cognitive levels, but then their instruction, their testing, and their grading tend to emphasize the lower behaviors levels, such as computation and comprehensions” (p. 650). NCTM (2013) states, “High expectations need to be publicly set for all students”. This is achieved through a focus on higher cognitive demand tasks. (NCTM, 2013) Wilson (1971) references “a vast change in the objectives of mathematics instruction in the 1960s” (p. 663). This change was “a shift from an emphasis on computation and application to an emphasis on comprehension and analysis” (p. 663). NCTM (2013) acknowledges this shift as well. “The exploration of developing intrinsic appreciation rather than extrinsic appreciation” (Wilson, 1971, p. 663) was included in this shift as well. Even though affective outcomes are harder to measure than cognitive outcomes, “attention must be paid to them in the evaluation of our instruction” (p. 663). “We want students to develop or maintain a strong positive feeling towards mathematics” (p. 663).

Wilson’s model is geared towards summative assessments, but could be adapted for formative assessments. Wilson also notes that the content he included reflects the secondary curriculum of the time that this model was published. The table of specifications “can be modified for a particular curriculum or classroom” (Wilson, 1971, p. 645). Lastly, Wilson mentions a need for multiple types of assessment. He states, “The evaluation of mathematics learning in terms of a single measure leads to incomplete or erroneous information” (p. 693). Wilson also comments on the lack of revolution in mathematics tests despite the revolution that has taken place in mathematics curricula. NCTM (2013) appears to still be waiting for this revolution forty-two years later.

Formative assessments are another evaluation tool. Many authors promote the use of both summative and formative assessments, while others prefer one to the other. Suurtamm (2012) believes that we should “move away from assessment as merely an event occurring at the end of a unit to provide a mark for reporting and move toward assessment as a process of ongoing feedback” (p. 28). Suurtamm argues that one way to achieve this is through the use of formative assessments. Formative assessments take on a variety of forms other than the stereotypical quizzes. “Observations, conferencing, questioning, progress reports, journal entries, and submission of problem-solving work” (p. 29) are all examples of formative assessments. Formative assessments not only “provide teachers with a window into students’ mathematical reasoning and sense making, [they also] create a forum for student and teacher discussion about such thinking” (p. 28). According to Black and Wiliam (1998) and Brookhart (2007) there appears to be a direct link between the use of formative assessments and raising academic achievement.

Hendrickson is an author that prefers formative assessments. Grouping high-stakes testing in with summative assessments, she suggests that the United States replace summative, high-stakes assessment with formative assessment. (Hendrickson, 2012, p. 489) Hendrickson justifies this statement by referencing Finland’s assessment practices. Finland “has scored first or second in mathematics during the past four administrations of the Programme for International Student Assessment (PISA)” (p. 488). The PISA is an international test administered every three years. It is administered to fifteen year-old students in fifty-seven countries, including the United States. In 2009, Finland’s average mathematics score was 541, while the United States had an average score of 487. (OECD, 2009) The United States mathematics score was reported as “statistically significantly below the average,” which was a score of 496. (OECD, 2009) “Kupiainen, Hautamaki, and Karjalainen (2009) suggest that Finland’s high scores on the PSA are due to the lack of high-stakes testing within the country”(Hendrickson, 2012, p. 488) and their focus on formative assessment. Finland’s use of formative assessment can be summed up into three main purposes.

“Purpose 1: Provide an adaptive environment to meet student needs.

Purpose 2: Provide informal information for teachers’ use in evaluating and providing feedback to students.

Purpose 3: Promote student self-evaluation within the classroom.”

(p.488)

Formative evaluations involve a lot of questioning, which “gives teachers access to student thinking” (p. 488) and insight into student misconceptions. This allows teachers to immediately identify where students are struggling and provide them with appropriate the instruction. (Hendrickson, 2012) I.e. an adaptive environment. The informal information gained by teachers is beneficial to both teachers and students. The constant influx of information allows teachers to adjust their instruction and cater it towards a specific group of students. Additionally, “teachers [are able to] comment on both students’ progress throughout the course and their mastery of the curriculum” (p. 488). In Finland, at the end of the school year, students are given “formative feedback and recommendations rather than a single grade” (p. 488). Students do not just receive a letter grade; they are given strategies for improvement. (Hendrickson, 2012) This feedback “allows students to be aware of their own thinking and develop the skills to evaluate their own work”(p. 488). If Finland is so adamant about formative assessments, why did they participate in the PISA? One school principal responded with, “some testing is thus ultimately necessary…if only to prove that regular testing is not” (p. 489).

Performance assessments are another evaluation tool and are often referred to as “alternative” assessments. The line between performance assessments and formative assessments can be a bit blurred at times as well. Performance assessments require some type of rubric and are usually considered to be less subjective than formative assessments. An assessment project conducted at Borel Middle School in San Mateo, California saw positive results when they began to incorporate performance assessments into their instructions. (Kitchen, Cherrington, Gates, Hitchings, Majka, Merk, & Truebow, 2002) The mathematics teachers at the school worked together to develop performance assessments that “aligned with the curriculum goals and objectives across the entire 6-8 curriculum” (p. 27). The teachers wanted to create assessments that “promoted higher-order thinking” and “require[d] students to apply their knowledge in real-life contexts” (p. 25). Four performance assessments were written for each grade level. The teachers dedicated three afternoons each quarter to grading these performance assessments. The amount of time they spent analyzing these performance assessment allowed them to pick up on student misconceptions. Additionally, the constant collaboration among teachers created consistency, in both instruction and assessment, from classroom to classroom. The teachers found that “student performance consistently improved during the first three years that the [assessments] were administered” (p. 26). The students also seemed to have a positive view of the assessments and often did extra work to receive a score of “exceeds-expectations”.

The STEM Model is a curriculum model geared towards performance assessments. However, it can be applied to summative assessments as well. (Billstein, 1998) The model is the product of the Six Through Eight Mathematics (STEM) project, which is one of the National Science Foundation’s Middle school mathematics curriculum projects. Billstein (1998) was the director of this project. The STEM curriculum is organized into modules and each module has a conceptual theme. (Billstein, 1998) The use of conceptual themes conveys mathematics as one big map, where all the topics are interconnected. (Billstein, 1998) Since “assessment criteria are not add-ons to the curriculum, but are [rather] an integral part of the teaching materials (p. 282), …STEM designed an assessment package on the basis of six principals” (p. 282). These principals sound very similar to NCTM’s Assessment Standards.

“The STEM Assessment Principals include:

1. Assessments should have the improvement of learning as its primary goal.
2. Documenting students’ achievements should be an integral part of the instructional process, not an add-on to it.
3. Assessment should include the active participation of students in open problems.
4. Assessment should reflect real-world applications.
5. Assessment should permit the full use of technology.
6. Assessment should use a variety of methods” (p. 283)

The STEM model focuses on open problems. “Open problems can be open-ended or open-response problems” (Billstein, 1998, p. 283). These problems have multiple solutions or multiple ways to obtain the solution. “Open questions usually revel more about students and their thinking than traditional questions” p. 283). Performance assessments usually involve an open problem, or a collection of open problems. In the STEM model, performance assessments are evaluated by a set of rubrics. One rubric is for the teacher to evaluate the student and the other is for the student to perform a self-assessment. (Billstein, 1998,) Like Finland, the STEM model stresses the importance of students’ self-assessment. Billstein states, “If a student’s self-assessment and the teacher’s assessment do not agree, a great opportunity arises for a discussion between the student and the teacher”(p. 294). These rubrics not only help provide consistency in a teacher’s evaluations, they also “clearly communicate to the students what is valued in the curriculum” (p. 284.) A “rubric [also] tells the student immediately where effort is needed to improve performance” (Billstein, 1998, p. 294).

Although there are differences in the assessment models described in this paper, there are similarities between them as well. One similarity is the continuous cycle embodied by all of the models. The models “use the information they acquire through the assessment process to make instructional decisions” (Moskal, 200, p. 193). Teacher’s are also able to “reflect on the thinking that [their] instruction may have engendered” (Kastberg, 2003, p. 403), which allows them to determine student misconceptions. This continuous cycle brings a teacher’s instruction and assessment into alignment. (Kastberg, 2003) The promotion of conceptual understanding is also evident in all of the models. This is even reflected in the assessment models geared towards summative assessments, which are usually thought to lack high cognitively demanding tasks. One underlying similarity is the models’ attention to consistency. When using an assessment model that deals with behaviors, it is important to be consistent in your classification and placement of behaviors. (Kastberg, 2003) When dealing with performance assessments it is important to have some type of rubric. The consistency that a model can help provide is key. However, even in models considered to be systematic, problems in consistency can still appear. For example, a student’s “response to an item may not illustrate the level of behavior intended” (Wilson, 1971, p. 664). Additionally, “the way a student solves a problem may determine the level of behavior” of the problem (p. 664). Ultimately, a teacher’s assessment should be “encouraging and supportive by nature and intended to improve learning” (Hendrickson, 2012, p. 488).

References

Billstein, R. (1998). Assessment: The STEM model. *Mathematics Teaching in the Middle*

*School 3*(4), 282-286, 294-296

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in*

*Education Principals, Policy, and Practices 5*(1), 7-74

Bloom, B. S., Hastings, J. T., & Madus, G. F. (1971). *Handbook on Formative and Summative*

*Evaluation of Student Learning*. McGraw Hill Inc.

Brookhart, S. M. (2007). Expanding views about formative classroom assessment: A

review of the literature. In J. H. McMillan (Ed.), *Formative Classroom Assessment:*

*Theory into Practice* (pp. 43-62). New York: Teachers College Press.

Carnegie Melon. (2013). *Formative vs. summative assessment*. Retrieved from

http://www.cmu.edu/teaching/assessment/howto/basics/formative-

summative.html

Hendrickson, K. A. (2012). Learning from Finland: Formative assessment. *The*

*Mathematics Teacehr 105*(7), 488-489.

Kitchen, R., Cherrington, A., Gates, J., Hitchings, J., Majka, M., Merk, M., & Trubow, G.

(2002). Supporting the reform through performance assessment. *Mathematics*

*Teachingin the Middle School 8*(1), 24-30

OECD. (2009). *PISA 2009 key findings*. Retrieved from

http://www.oecd.org/pisa/pisa2009keyfindings.htm

Moskal, B. M. (2000). An assessment model for the mathematics classroom. *Mathematics*

*Teaching in the Middle School 6*(3), 192-194

National Council of Teachers of Mathematics. (2013). *Assessment standards for school*

*Mathematics*. Retrieved from

http://www.nctm.org/standards/content.aspx?id=24079

Suurtamm, C. (2012). Assessment can support reasoning and sense making. *The*

*Mathematics Teacher 106*(1), 28-33

Wilson, J. W. (1971). Evaluation of learning in secondary school mathematics.

In Bloom, B. S., Hastings, J. T., & Madus, G. F., *Handbook on Formative and Summative*

*Evaluation of Student Learning* (pp. 642-696). McGraw Hill Inc.