rDOK Procedure for Classifying the Cognitive Complexity of Mathematics Items

By: Alexander Hoffman
David Whitcomb

ALEDEV CONSULTING

The RTD (Rigorous Test Development) project is an attempt to build a professionalized content development practice that focuses on individual item quality, particularly by leaning into the importance of validity throughout the content development process. It assumes that content development professionals develop professional judgment that can be raised, honed and calibrated by providing frameworks and clarifying expectations in ways that account for the constraints and demands of typical practice within test development, today. RTD is a conscious and deliberate attempt to respond to the disparity in status, training and shared knowledgebases between psychometrically oriented professionals and content development professionals.

Table of Contents

rDOK Procedure for Mathematics

The document *RTD Approach to Using Norman Webb's Depth of Knowledge (DOK) Typology of Cognitive Complexity* lays out a general approach understanding DOK and general approaches for applying it to standards and for applying it to items. However, actual application to items varies a bit from content area to content area. This is simply a product of the nature of the different disciplines and the assessments of them. For example, Mathematics items tend to focus on math problems (i.e., recognizing, defining, finding entry points and solution paths, implementing those paths and arriving at solutions for problems), whereas there is no such as an ELA (English Language Arts) *problem*. ELA is instead focused on communication – both understanding that of others and expressing oneself – including mechanical skills but with greater emphasis on broader thinking skills. There *are* areas of overlap, but – as befits entirely different content areas – the fundamental nature of items is usually quite different. This requires slightly different approaches when *applying* DOK.

DOK Basics (Mathematics)

Webb's Depth of Knowledge typology has four levels (see table). These levels can be explained specifically in the context of addressing the cognitive paths taken by students and test takers in response to math problems, tasks and items.

DOK 1 applications rely on *automaticity*, a major goal of math instruction and practice for students. Mathematics is often built around the development of mathematical toolbox and students have achieved mastery when they quickly recognize the appropriate tool and can apply it with practiced ease. This may include care in following the steps of an algorithm, but the steps are well known and do not rely on deliberation about which step to take next or how to apply it. Even a multi-step solution path can be DOK 1 when the steps are well-known in advance, and thus solving the problem is the simple walking down a well-trod path.

DOK 2 applications add some amount *tactical* conscious deliberation to the application of skills and/or concepts. For example, the test takers might have to make deliberative decisions about which values to plug into a formula. Similarly, the test taker might have to put some conscious decision-making into interpreting or translating something presented to them – be it a narrative, a graph or an equation. Addressing a multi-step problem can be DOK 2 when the test taker has to take stock at particular steps, considering of what to do next when each decision is fairly evident once deliberatively considered.

Most DOK 2 applications can become DOK 1 applications when they are sufficiently practiced that they becomes rote and automatic. Conversely, skills and concepts who application should be rote or automatic may instead be DOK 2 application for test takers who have not yet achieved automaticity. For example, many students can immediately recognize that 49 or 64 are squares of 7 and 8, but others may still have to work through an uncertain process to figure that out.

rDOK Procedure for Mathematics

Level	Name	Description
DOK 1	Recall	Rote use of algorithms or procedures. Direct application of definitions, such as recognizing examples or simple classification into previously well-known categories. Absence of conscious and deliberative decision-making, making for skilled automaticity.
DOK 2	Skill/ Concept	Basic decision making. Applications of known skills that require conscious and deliberative decisions, such as about what values to plug into a formula or selecting an appropriate tool or algorithm. Includes transforming from one form to another (e.g., from word problem into algebraic or arithmetic expressions) when it requires conscious and deliberative effort (i.e., beyond the automaticity of DOK 1).
DOK 3	Strategic Thinking	Metacognition about the cognitive path of solving a problem. Planning a cognitive path, monitoring and evaluating the progress and/or providing explanations of decisions/thinking process. Conscious and deliberative decisions about what approach to take.
DOK 4	Extended Thinking	Thinking that is extended across multiple contexts or concerns in ways that connect those contexts or concerns. Arriving at generalizations based upon a range of information or ideas. Analysis that includes multiple factors or issues and account for those issues in the final product.

DOK 3 applications are about *strategic* reasoning and metacognitive thinking about the solution path. When this is done before embarking down the path, it takes the form of serious planning based on recognizing intermediate goals that must be reached and then figuring out how to get to them—in order to reach the end solution. When done after completing the solution path, it may take the form of explaining the reasoning behind each step. When done alongside the work, it may take the form of the test takers explaining that each decision (e.g., as in a mathematical proof) that they made. However, the most common DOK 3 application may simply be monitoring and/or evaluating one's progress so that one may change course if necessary (see discussion of SMP 1, below).

DOK 4 applications simply do not appear on timed, on-demand, standards-based standardized assessment.

A Range Cognitive Complexity for the Range of Typical Test Takers

As explained in *RTD Approach to DOK*, one must recognize that the cognitive complexity of the work done by a range of students/test takers can vary. The most skilled and proficient test takers may work with the automaticity of DOK 1, while less proficient students may work their way through to the correct response through DOK 2 applications of the same skills and/or concepts. Similarly, less proficient students may have to carefully survey potential solution paths in order to select a successful approach or carefully and consciously monitor their progress (i.e., DOK 3), whereas more proficient test takers can simply plunge into a novel task, trusting that they will quickly spot the correct next step when they need to (i.e., DOK 2).

Those evaluating the cognitive complexity of items should be careful that they not simply evaluate *their own* paths through items – either as adults or based upon their projections of what their paths might have been at the appropriate age. Yes, they can make use of those evaluations, but they should also consciously consider the nature of both more and of less proficient test takers' solution paths when determining the minimum and maximum DOK classifications for items.

In fact, many items *do* have the same minimum and maximum DOK applications. Problems that would be somewhat novel for virtually *any* test taker are often DOK 2 for all test takers. Problems that should be solved via a well-established algorithm and would not be reasonable to test takers to see another solution path in the context of the assessment will be DOK 1 for all test takers. However, problems and items whose cognitive path varies in ways explained above can have different minimum and maximum DOK applications, for more and less proficient test takers, respectively.

Focus on Targeted Cognition

As explained in *RTD Approach to DOK*, all items included in timed, on-demand, standards-based standardized assessments are attempted in the context of a DOK 3 task. That is, test takers have to manage their time, decide when to bail on an item, decide the conditions under which they will return to check their work or to reattempt to complete an item. Thus, because the context of any cognitive work on a such an assessment is always a larger DOK 3 task, the context of the application does not determine the DOK classification of an item. This even includes the cognitive complexity of just the individual item—as that would still be an arbitrary line to draw.

Instead, the DOK classification of item should be *based upon the application of the targeted cognition.* Other aspects of the item and the solution path clearly are important, and may figure greatly in the *difficulty* of the item, but the purpose of the classifying the cognitive complexity of items is to ensure that they are assessed at levels prescribed in the standards. For example, one cannot compensate for a reduction in the complexity of the application of a skill/concept in a standard by embedding it in an item for which it is difficult to find an entry point. An item that is presented in an unclear or ambiguous manner may demand more complex cognition for the test taker to even make sense of, but

that does not compensate for presenting a DOK 2 standard in a way that merely requires a memorized response from test takers.

Therefore, it is critically important to recognize which part(s) or step(s) of the solution path entail the targeted cognition (i.e., skills or concepts from the aligned standard(s)). For example, when a standard is specifically about words problems or applying a skill in real world context, making sense of the problem is, in fact, part of the targeted cognition. Evidence of proficiency with such standards require students to engage in some degree of interpretation and sense-making. On the other hand, standards that simply lay out a mathematical tool or concept do *not* require such skills when demonstrating proficiency, and their inclusions as part of the solution path should *not* factor into DOK classification.

SMP 1: Make Sense of Problems and Persevere in Solving Them

The very first Standard for Mathematical Practice is quite relevant to various aspects of DOK classification. These absolutely critical and broadly applicable habits are among the most important lessons that students can learn. They truly are lessons worth learning for a lifetime and students should use them every day. However, understanding their application can make DOK classification a bit more difficult.

As explained above, making sense of problems and finding an entry point are often built into learning/curricular standards and even when it is *not* built into a standard, they may still be required to respond to an item. Items may be aligned to content standards that *are* built around these two skills (i.e., finding entry points and/or making sense of problems). Examples of the former will usually include at least DOK level 2, as the need to deliberatively make sense of something is the opposite of the automaticity that typifies DOK 1 cognition. Items aligned to content standards that are *not* built around these two skills often include DOK 1 cognition, as when the application of the actual targeted cognition described in the standard is rote or automatic – at least for some test takers.

This becomes even more complicated with regard to *persevering in solving problems.* SMP 1's explanation includes, "They monitor and evaluate their progress and change course if necessary." Of course, students should do this. And any student can make a mistake on virtually any item, recognize that something is wrong and then go back and fix it. This kind of metacognitive awareness of progress towards a solution *is* DOK 3 cognition. However, as with *making sense of items,* one should only factor that into the classification *of the item* when the targeted cognition/aligned standard calls for that skill/habit. Otherwise, *all* items would have maximum DOK of at least level 3, rendering the entire concept of maximum DOK level useless and uninformative.

Be Careful of Masquerading Items

Honestly, the purpose of recognizing the cognitive complexity of items—and of the applications of the targeted cognition within test takers' solution paths—is to do a better job of including items that required test takers to engage in levels of cognitive complexity that are as great as those described or assumed in the standards. There is a well-known

history of dumbing down or simplifying the appearance of standards on standardized tests. Usually, this is entirely unintentional and is simply a product of the various constraints of standardized tests and their development. Everyone wants items and assessments that well reflect the contents and complexity of the standards.

Unfortunately, standardized tests' frequent reliance on the multiple choice item format often can lead to unexpected cognitive complexity levels of items. For example, back solving (i.e., plugging answer options back into a problem to see which one(s) work) is usually a DOK 1 approach, even for items whose *intended* solutions paths are DOK levels 2 or 3. Thus, many problems that would be DOK levels 2 or 3 as constructed response items have a minimum DOK at level 1. Other tasks can actually call on greater cognitive complexity when presented in a multiple choice format, though this is far less common. Graphing the line of an equation is often a DOK 1 task, but selecting the graph that shows the correct line for an equation can call on a level of analysis that is more complex, and when not a rote skill rises to DOK 2.

Therefore, it is vital that items be evaluated *as presented*, which means thinking through how test takers will respond to them *as presented*. One must read through the whole item—including the answer options—and consider the impact of the particular answer options when classifying the minimum and maximum DOK level of an item.

This is quite commonly a problem with items that are aligned to DOK 3 standards. Such cognition usually entails voluntarily exercising a habit of metacognition, rather than simply following a prompted path. For example, many mathematics standards requires students to offer or explain their reasoning and/or explain *why* something is true. These *standards* assume constructed response tasks and do not address the cognitively simpler task of recognizing valid reasons when presented with them. This simpler task may entail simply evaluating the truth/falseness of an offered reason or evaluating its relevance to another statement. These are often DOK 2 applications of the targeted cognition, rather than the DOK 3 application anticipated by a standard. Inclusion of the word "because" is one clear indicator this issue may have appeared.

The kind of deliberative decision making that is important to DOK 2 or DOK 3 applications relies on a certain amount of novelty in the problem faced. When students have practiced a particular type of problem sufficiently, they can become sufficiently proficient that they are able to work at a lower level of cognitive demand. This is often a *good* thing, but it does lower the DOK level of their work. As curriculum and assessments becomes more and more aligned, this kind of overlap can become more common. If standardized assessments becomes sufficiently predictable, students can be taught to the test such that the novelty of problems is lost, for them. Thus, the DOK level of items can depend on the enacted curriculum for students. DOK classification must take this into account.

Poorly Written Items

Frankly, poorly written items add additional cognitive burdens on test takers. They can turn simple applications of mathematical concepts and skills into more complex efforts at interpretation, and can also make items needlessly difficult. Sometimes, items

should require test takers to do work to figure out what is relevant and what they need to do, but sometimes this is additional work that simply creates unnecessary barriers to test takers' exhibitions of their abilities.

There is no question that poorly written items present additional challenges to those who would classify their cognitive complexity. The work of correctly responding to such items can include more complex cognition than the aligned standard describes or assumes, but the DOK classification of the item must nonetheless focus on *the application of the targeted cognition*, as explained above.

This is inevitably frustrating to committed and caring professionals when they recognize that that has occurred—particularly when DOK classification is done too late to change an item (e.g., after field testing). On the other hand, recognizing this issue earlier in the process can help to flag items that need more work before they go to field testing.

When doing the work of final or confirming DOK classification, the assessment professional must assume that the test taker correctly understood the intent of the item. Otherwise, something other than their application of the targeted cognition becomes the object of their classification work.

Poorly Aligned Items

There may be items included in a DOK classification project that are, shall we say, poorly aligned to their targeted standard(s). If these items cannot be modified or pulled, they still need to be classified for their minimum and maximum DOK levels. However, poorly aligned items may not require the targeted cognition (i.e., the KSAs include in the standard) and/or may have *Additional KSAs* that are greater barriers to reaching a correct response than those that make up the standard.

RTD recognizes that they *Key KSAs* of an item are those KSAs that differentiate successful test takers from unsuccessful test takers. That is, the KSAs that successful test takers are able to use but that unsuccessful test takers are not. These are the KSAs that any item actually elicits evidence of. Poorly aligned items are those for which the *Key KSAs* are *not* part of the purported aligned standard.

In these cases, classifiers should consider the cognitive complexity of the application of those *Key KSAs* and *not* focus on the KSAs that make up the purportedly aligned standard. The sole—but perhaps too common—exception to this is when it is SMP1 *making sense of problems* and its reading and interpretations skills. *Those* KSAs can be ignored when classifying items for DOK (unless the item is explicitly aligned to SMP 1).

DOK is not Difficulty

Every single one of us sometimes finds ourself confusing item difficulty and DOK for a moment. There are items whose difficulty (or ease) seem a little at odds with their DOK level, and it can take a moment to overcome an initial judgment that was distracted by that fact. Deliberate care and thoughtfulness when engaging in this work can take care of that.

Math Item DOK Classification Procedure

Each item should be evaluated against *each* DOK level, thus allowing for each item to be classified with up to three recognized DOK levels. This recognition that different solutions paths can have different levels of cognitive complexity *does* stand in contrast with traditional consideration of cognitive complexity.

An items should be considered to elicit a particular level of cognitive complexity when a significant share of test takers might apply the targeted cognition at that DOK level.

DOK 1

- Some test takers may respond simply by using well-practiced algorithms and approaches, such that the *entire* solution path is quickly apparent. This may include:
  - Cognitively simple recognition and recall.
  - Direct application of definitions.
  - Other problems whose like test takers have seen before.
- Some test takers may select their response by use of back solving.

DOK 2

- Test takers may respond using a well-established mathematical tool or algorithm, but there remain decisions about how to use the tool whose answer may not be immediately apparent to some test takers. Such decisions must have the potential to—if made poorly—fail to lead to the correct answer.
- Some test takers may respond with a solution path whose each step is obvious as the previous step is completed, but were not clear *before* the first step was completed.
- Some test takers may have to translate or transform the problem in order recognize the nature of the problem and only then can select the correct algorithmic response.
- Some test takers may apply simple reasoning and thereby *interpret* the meaning of a result and/or draw a conclusion.
- Some test takers may respond with an algorithmic solution path, but for some students this path is not sufficiently well known and practiced as to constitute a *rote* response.

DOK 3

- Problems that are sufficiently novel to some test takers that they present multiple truly different potential solutions paths to choose from and none of them is immediately obviously sure to work successfully.
- Some test takers must select a solution path from among multiple considered paths by judging which is most likely to prove successful, without being sure that it will work.
- Some test takers may plan out a solution path up front, for themselves.
- Items which require test takers to explain their reasoning
- Items for which some students may monitor and evaluate their progress towards a solution and restart when appropriate, *and whose aligned standard calls on such a skill.*