What is RTD's Radical Empathy

By: Marjorie Wine
        Alexander Hoffman

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

Defining Radical Empathy?

Radical Empathy is a rigorous and disciplined technique used by CDPs (content development professionals) to examine items through the eyes of *the range of typical test takers* for an assessment. This difficult practice is important to assuring that items *elicit evidence of the targeted cognition for the range of typical test takers*. It is the heart of the RTD (Rigorous Test Development) Item Alignment Examination procedure.

Radical Empathy has nothing to do with sympathy for test takers. That is, it does not call on CDPs to engage their own emotions in reaction to test takers, their experiences or their challenges. Rather, it requires CDPs to be able to put aside their emotions, reactions and even their own thinking in response to an item so that they can put on the hat, step into the shoes and look through the eyes of a test taker who is unlike themself. When engaged in Radical Empathy, a CDP must try to think through the item as that hypothetical test taker would.

While that level of empathy is usual, what is radical about Radical Empathy is the breadth of its application. That is, a single item should be considered through as many perspectives as possible, so long as those perspectives can yield a different cognitive path (see *Metacognition,* below) through an item from the stimulus all the way to providing or selecting an answer. Test takers can vary in the strategies they adopt, in their backgrounds and identities, in their command both of the Targeted Cognition and of additional KSAs (knowledge, skills and/or abilities). Any of these kinds of variation can lead to a different path through an item. Of course, even setting those differences aside, there are can still be multiple paths through an item that all rely on full command of the content and for which these other factors do not play a role (see *Awareness of Multiple Paths*, below).

The goal of Radical Empathy is to uncover and raise awareness of those paths so that individual items can do a better with a wider range of test takers.

About This Document

This document serves as an introduction to the practice of Radical Empathy. It explains what it is, why it is important and the basic path to developing proficiency with this practice. However, it is *not* a how-to, lesson plan or outline of a curriculum.

Radical Empathy is not a procedure and it cannot be learned simply by following step-by-step directions. Rather, it is practice that must be developed over time. Learning to engage in this practice requires careful thought, deep reflection, patience and openness to real growth. Watching others sharing their practice of Radical Empathy transparently can be quite helpful. Getting feedback and direction on one's own attempts can be just as helpful. In other words, it is a demanding practice that benefits from teaching and mentorship and is difficult to teach oneself.

Thus, we hope that this document may inspire current and aspiring CDPs to engage in that difficult learning because they see the potential value of this practice. We hope that

this document explains the practice and what it requires sufficiently for them (you) to embark on that learning path.

Of course, we hope to provide more resources and opportunities in the future to support CDP's professional development.

The Value of Radical Empathy in Item Development

Items are the basic building blocks of any assessment. If the evidence supplied by tests and the inferences made upon them is to be valid, items themselves must *elicit evidence of the targeted cognition for the range of typical test takers*. This is what we call *item validity.* While there are myriad content issues and item design issues that can contribute to – or detract from – item validity, perhaps the greatest challenge to producing high quality tests is the range of test takers for which they must be valid.

When items are not valid, the evidence they produce cannot be trusted for all test takers. Instead, items lead to false positive and false negative inferences. That is, some test takers answer items correctly without proficiency with the targeted cognition, and some answer items incorrectly despite possessing proficiency with the targeted cognition. A standardized assessment that yields many false positive (Type I errors) and false negatives (Type II errors) is far less useful to anyone. An assessment that performs less accurately for some groups of test takers than for others is not only less useful, it might be immoral – or even illegal. The is bias. This can produce illegal discrimination in education and employment opportunities by race, gender and other rightfully protected statuses.

Radical Empathy is, in our view, the best approach for addressing the issue of false positives and false negative results across the *range* typical test takers. Of course, those content and item design considerations must also be considered (see *The Role of Content Knowledge*, below), and Radical Empathy is a practice that allows them to considered across that range of test takers. This is how a CDP can determine whether the stem plus the key (i.e., the correct answer) might yield false positive evidence and the stem plus distractors (i.e., the incorrect answer options) might yield false negative answers.

Without Radical Empathy, test developers all too often offer tests that work for test takers like themselves, with their kinds of backgrounds and thinking. It is not simply that people unlike themselves will fail those tests, but rather it is too easy for some of them to *undeservedly* fail and for others to *undeservedly* pass. Their scores simply will not reliable – or, in this case, support valid inferences – for those who depend upon them.

When to Make Use of Radical Empathy

We really dove into thinking about Radical Empathy and what it requires of CDPs in the context of RTD's IAE (Item Alignment Examination procedure). Obviously, Radical Empathy is key to IAE. However, the full IAE procedure is very demanding and is often more than is needed at most points in the item development process. Radical Empathy, on the other hand, is almost always an appropriate tool. It should always be available, and should become automatic in short order for experienced CDPs.

The CDP problem of viewing items just through one's own eyes or through the eyes of a remembered self never quite goes away; there is no point when that view is sufficient. Any time CDPs view an item and are in a position to edit it or to give feedback on it, they should try to view item through the eyes of hypothetical test takers unlike themselves.

Thus, the question is *not* when to make use of Radical Empathy, but rather *how broadly* to repeat it. How many different readings of the items might there be? How many authentic cognitive paths through the item should a CDP look for?

In practice, this becomes a matter of professional judgment. Experienced CDPs who are skilled with Radical Empathy likely will have internalized it and will usually notice features of items that may prompt different cognitive paths for different test takers – both different in backgrounds and different in approaches to the item. Having noticed those features, they then should think through those test takers' cognitive path(s) to develop their own understanding of how different test takers will respond to the item. Even such experienced CDPs may also decide to quickly run through checks of additional perspectives to make sure they are not missing something.

Less experienced and proficient CDPs will have to do what less proficient people do all that time. That is, they will have to be more conscious and intentional in their application of a process – in this case, Radical Empathy – and because of that simply will lack the time to consider as many perspectives. However, with experience they will internalize the practice and get better at noticing when to put on different personae.

A Path to Radical Empathy

There is a clear path – or progression – in the development of one's ability to engage in Radical Empathy. There may be others as well, but this is one path that we think is most common and accessible to content development professionals.

At the most basic level, a potential CDP must understand items in a content area and be able to respond to them.

At the second level, a potential CDP must be able to explain how they responded to the item. This might be as simple as showing one's work or explaining the major strategic decisions that they made when responding to the item.

At the third level, a potential CDP must be able to consciously notice the elements in an item that induced them to take the cognitive path they did when responding to the item. For example, they might say, "Because stem asked for an example of [X], I went back to the passage and looked for the word [blah]. I knew that I would find the answer around that," or "the phrase 'left over' in the question signaled that I needed to subtract one number from the other."

At the fourth level, a CDP must be able engage in *close reading* of items. That is, they must be able to consciously notice elements in items that might lead *other* test takers to take different paths than they themselves did. For example, "I can see where some students might be confused by this part of the instructions. If they interpret them like [M] then I think that they will end up doing [X], but if they interpret them like [N] then I think they

might end up doing [Y]," or "I don't think that kids from rural areas will know what [Z] means."

At the fifth level, a CDP must be hyper aware of every little step, thought and decision along the cognitive that that *they* took when responding to an item. For example, "…$19.74. ha. I was born in 1974. I notice that every time. I then…." This level of awareness might be the first big jump from typical practice and the first step that less experienced CDPs need to think about and work on to move closer to using Radical Empathy in their work.

The sixth level parallels the fifth level, rather than building up on it. Here, CDPs must deeply understand that there are multiple cognitive paths through most any item. They must understand that even among very similar test takers who all arrive at the same answer, there can be multiple paths to get there that tap into different KSA – both from the Targeted Cognition and from additional KSAs.

The final level is the practice of Radical Empathy. At this level, CDPs consider multiple perspectives with the level of details and richness that they recognized in their own cognitive path. Over time, they should get better at doing this with a broader and broader range of perspectives.

It is impossible to expect that anyone could really understand the range of all potential test takers. There is no point at which any CDP could understand *every* possible cognitive path through an item. No matter what, items should be reviewed by a diverse range of people (e.g., in review panels, among CDPs). There simply is no resting point at which a CDP can say, "Yeah, I've got nothing more to learn about test taker perspectives."

Close Reading of Items

As suggested above, there is no way to engage in CDP work at a professional level without being able to read items closely. That is, CDPs *must* read items beyond their own experience with them, beyond how they solved the items, beyond how they themselves understood the items. They must be able to step back from the item and examine it from a more removed and objective perspective.

In practice, we often recommend first going through the item as oneself, just responding to the item *at speed*. That is, just do it the way you would do it as yourself, as fast/slow and as sloppy/careful as *you* would. Some do this as their adult current selves (i.e., what we usually recommend) and others always put on the persona of their younger selves (i.e., at the age of the intended test takers for the particular item). It is hard to be open to others' (imagined) experiences with an item if one has not given one's own views a chance to come out. This is simply the first step in taking in an item.

The close reading of the item them comes out when one is able to step a bit outside one's initial experience of the item and is able to examine and critique the item. This can include noticing features that prompted one's owns choices and features that might be understood differently by others. It might include features that one just ignored – rightly *or* wrongly – but others would notice and process differently.

At this point, a CDP might notice elements that would be more comprehensible to some takers than others. They might notice features that virtually any test taker could be confused by. They might notice things that are *missing* from items, but they savvily took for granted when they originally read the item.

Close reading is about a) noticing what is in the item (and what is not in the item) as it as it would presented to potential test takers and b) noticing elements and features that contribute to the decisions that test taker will make when they have their opportunities to respond to the item. These can prompt things as obvious as conscious and strategic decisions and down to things as subtle as unconscious minor distractions.

Metacognition

Metacognition and metacognitive awareness to the degree that Radical Empathy requires truly is an enormous leap from historical practice. It rests on putting aside the assumption that people necessarily think in a linear way, maintain perfect focus throughout a task, take only rational steps and that thinking is clean and clear.

Instead, human thinking – even by real experts – can be incredibly non-linear. People get distracted and come back to their main thought, they often make unsupported jumps in their thinking and depend enormously on their unconscious mind for associations and shortcut. Honestly, any time I see my birth year as a number in a problem, I notice it. (1974 is actually the birth year of *one of our brothers*, not either of ours, though). That does not keep me from correctly responding to situation in front of me. However, if I were honest about my thinking process through a math problem, I would have to acknowledge that minor little bump along *my* cognitive path.

People often engage in close reading of items with a simplifying assumption that people take in phrases and sentences all at once, like we sightread words without consciously working through the letters together in order. In fact, we do *not* withhold judgement of the meaning of sentence until we get to the end, and we *can* get wrong ideas before we have to correct them. This is simply the normal task of constructing meaning when reading – or even when listening to someone else speak. We bring our own experiences to bear and interpret as things unfold. Little things therefore can distract us. For example, every time I hear someone say "Fellas," I think – and sometimes mutter – "Yeah?" Our minds are messy, can be fidgety and do *not* work like computers do.

Our minds are *messy*.

This application of metacognition is about accepting all that messiness. It is about exploring our own thinking with clear eyes and respect, because that is how it really happened.

...then I saw "brownies" and I just stopped for like half a second to picture a thick warm dark – not too dense – brownie and then went on...

...17 plus 8 equals 24nothat'snotright. It equals 25....

...she mentioned a front door and I saw in my head the doorknocker on the front door of the house I grew up in...

...before I even read the stem, I noticed the word "best" set in bold type and just got annoyed...

...When I saw 5 and 12 in a math item, I thought "That's a 5-12-13 right triangle; the answer's 13" and jumped to the answer options. I went back to make sure, but I wasn't really looking that closely any more...

This kind of metacognition is largely a replacement for what is described above as the second level of reading an item. Like that second level, CDPs do this *after* having already gone through the item authentically as themselves. Like the second level, this is a reflective step, looking back on how they responded to the item. Only in this form, the results are *not* what they would want to show their teacher or share with their student. Rather, it acknowledges the messiness of their thinking. The sheer normal human messiness.

Reflecting like this on one's own cognitive path through items is *enormously* important because this parallels the authentic thinking that test takers will engage in. This is where a CDP can perhaps spot *why* they made the conscious and unconscious decisions that they did. This is how a CDP gets a better view of exactly what KSAs they used when responding to the item, and which ones were helpful to them getting to their answer.

Just as importantly, this kind of examination shows a CDP where they might have gone wrong, but did not. This shows the mistakes that the CDP was able to recover from. This is like making math students show their work and requiring them to do it in pen so that they cannot erase the stuff that did not work out – and not allowing them to do any mental math.

Why does this matter? Because this is the approach CDPs need to use to understand how *others'* cognitive paths might go. Understanding others' potential paths starts with really understanding one's own. Imagining the thinking of another person requires really acknowledging the complexity and messiness of one's own. If CDPs imagine that test takers will engage in idealized and linear cognitive paths the simply will not be able to develop tests that measure their actual proficiencies with cognitive standards.

Awareness of Multiple Paths

The second big leap towards being able to use Radical Empathy can be taken in parallel with developing the ability to engage in metacognition, because it does not build upon it. Metacognition is inwardly focused, while awareness of multiple paths is focused on others.

As a starting point, imagine the following math item, which is *very* similar to an item we recently looked at with a group a very experienced CDPs.

Jeremy and his brother went to the store to buy ingredients for lunch. They decided that they wanted to make peanut butter and jelly sandwiches all week. A loaf of bread cost $2.95, crunchy peanut butter cost $3.45 and the raspberry jelly cost $4.50.

Jeremy had $24.35 when they got to the store. How much money did they have left after they made their purchases?

a) $7.95
b) $10.90
c) $13.45
d) $35.95

All five of us arrived at the correct answer, but we each took a different approach.

- One person rounded everything in their head. *~3 + ~3 + ~4 = ~10. $24 - ~10 = ~$14. Only one answer is near $14. C.*

- One person did it the straightforward and diligent way. They took out a piece of paper and did the arithmetic the long way. *24.35 - 2.95 = 21.40. 21.40 - 3.45 = $17.95. 17.95 – 4.50 = 13.45. $13.45. C.*

- One person tried a slightly more sophisticated approach in their head. They added up all the costs *(2.95 + 3.45 + 4.50 = 10.90)* before subtracting that total from the $24.35. They came to $14.45, but they already felt like they had grouped/borrowed incorrectly before even looking through the answer options, so they pulled out a piece of paper and wrote everything down. They still got 10.90. Then, they subtracted on paper. *24.35 - 10.90 = 13.45. $13.45. C*

- The fourth person did not want to subtract. They added up the costs (*2.95 + 3.45 + 4.50 = 10.90*) in their head and then went through the answer options, adding each one to that subtotal to see what would result *in $24.35. 10.90 + 9.95 = 19, way too low. 10.90 + no, wait, that's what I already have, so that's not it. 10.90 + 13.45 = 24.35. $24.35. C.* Done. But they wanted to make sure so they looked at d (35.95) and ruled that out as obviously being far too big.

- The last person added up all the costs (*2.95 + 3.45 + 4.50 = 10.90)* in their head and then subtracted that from 23.35 and arrived (correctly) at *13.45. $13.45. C.*

Five different approaches. Each equally valid. Estimate. Just subtract each in order. Use mental math skills to add and then subtract, but double check it on paper. Skip subtraction and just back solve. Just do it all precisely in your head *without* doublechecking your work. None of us selected a bad method. All of us arrived at the right answer.

This is a *very* simple math problem. If it is not a first grade problem, it is a second grade problem. But we each approached the problem differently. Without any stumbles on content, on context or on background, without any of us being thrown or distracted by anything, without any of adopting a problematic strategy, we came up with five different paths through the problem. Five

white educators with graduate degrees who have lived their whole lives in this country came up with five different approaches – and the least math confident of the group jumped to the most careful and diligent approach.

Imagine how many different paths there are through this item. We can think of at least a handful more, even without considering students who might be the least bit confused (e.g., a test taker who whose command of English is poor and does not know what a sandwich or peanut butter is), distracted (e.g., a test taker who did not get breakfast and is hungry for lunch) or actually has trouble with math. Consider how many ways a test taker could make a mistake when trying to set up the arithmetic from this word problem. And then there are the borrowing and carrying issues.

People who can do items fairly easily do not understand how differently other people might approach them. It is a big leap to realize how many different cognitive paths test takers might take in response to an item. It is not *just* about all of the potential mistakes and unfruitful paths – though those are *enormously* important. Even with simple items, and even when considering only paths that lead to the correct answer, there are usually more paths than are obvious.

More complex items offer even more potential cognitive paths. When one considers how lived experiences and different background knowledge can shape test takers understanding of items? When one considers how test takers with different degrees of skill or comfort might respond to an item? When one considers how different personalities might respond to an item? When one considered how different test taking strategies or approaches to problem solving can shape how test takers respond to items?

The example above offers just a tiny glimpse to the range of typical test takers.

Radical Empathy: Applying Metacognition to the Range of Typical Test Takers

In this RTD context, the practice of *empathy* is to imagine the cognitive path of a test taker unlike oneself – and doing so at level of messy detail like the practice of metacognition. This is difficult and disciplined work that – like professional practices in so many fields – takes years of experience to learn to do well. This kind of imaginative empathy is very demanding on a CDP, requiring that they do *more* than just put on the hat, step into the shoes and look through the eyes of a another human being. When truly mastered, Radical Empathy is quite like method acting. CDPs not only develop a sense of another's perspective, but actually *adopt it* for long enough think through it with that metacognitive level of awareness that is difficult enough to do with one's own thinking.

This is not simply an exercise in fiction writing. Rather, it akin to writing from someone else's perspective in a way that feels true to people more like the character than like the author. It relies on understanding other people *and how they think* in a way that recognizes who they really are. It is not about condescending to them, lionizing them or recognizing only surface differences. Yes, it is an imaginative act, but it is about searching for real truth.

Applying empathy so consciously like this items leans on close reading of items. A CDP must be aware of all the tiny details of items that that a very different person might notice and react to differently than the CDP did themselves. It requires the CDP to follow

that *other* persona's path, rather than sticking to their own. This is simultaneously highly *creative* work and highly *disciplined* work. It requires a creativity that is as directed as that of engineer trying to solve a difficult problem, but is applied to challenge of understanding another human being – though in a very limited context.

CDPs can get better at this, over time. Certainly, they can lean on their own improving metacognitive awareness of their own thinking. They can lean on their experiences with other people – including people who are quite different than themselves. They can learn about people indirectly, perhaps by listening to others explain their understanding and experiences, perhaps by reading scholarly works, new reports and even literature.

This act of imaginative empathy becomes *radical* when it is applied repeatedly and broadly. CDPs must get better at this kind conscious examination of hypothetical test takers' cognitive paths through items, and they must try to develop their ability to do this with more and more of the range of typical test takers. This includes test takers from different demographic groups, test takers who apply different strategies to responding to items, test takers with different levels of command of the KSAs in the standard and of other KSAs and even test takers whose physical, emotional and cognitive abilities might be very different than those of the CDP. (See our *RTD Radical Empathy Personae* for simple explanations of a few dozen kinds of differences that a CDP might consider.)

As we explained above, the practice of Radical Empathy requires CDPs to develop professional judgment to let them know which sorts of differences to consider (e.g., to notice when gender differences might suggest different cognitive paths in response to an item) and when to stop looking for more. As we explained above, content development requires groups or teams of people to really refine items well. This is because no single individual could possibly understand the entire range of typical test takers. External Review committees, therefore, are critical. Even when their members lack real proficiency with the practice of Radical Empathy, their different backgrounds, experience and relationships to the world – and even to the content – allow them to see issues that some test takers had encounter that would be quite difficult of a particular CDP to see. Furthermore, by listening to these panelist and asking then to explain even more, a CDP can work on better understanding some of *those* perspectives.

The Role of Content Knowledge

While the practice of Radical Empathy applies across all content areas and all kinds of test takers – from small children in elementary school to highly educated aspiring professional taking professional licensure exams – a CDP cannot simply apply it themselves across content areas. Content knowledge is very important to effective application of Radical Empathy.

For example, not having gone to medical school, we cannot anticipate or imagine how their graduates might respond to items on medical board exams. We generally lack the knowledge to even understand the questions, let alone to imagine the kinds of

mistakes that these test takers might make as they attempt to respond to them. We do not know what confusions might arise, what common errors might be or what the easy associations are. (In fact, this makes the development of valid medical board exam items particularly challenging, because many of the assessment professionals who work on them lack the content expertise and the SME (subject matter experts) who contribute to them lack the assessment expertise (e.g., skill with Radical Empathy).)

This issue is not limited to high level technical content. CDPs must know their content areas well enough to anticipate the types of problems, issues, distractions and confusions that test takers might have. That requires understanding the perspectives of learners of *that* content areas at *that* range of learning. CDPs may come to the professional with some of that understanding in hand, and then develop it further through their CDP careers. However, this is often easier to expand across ages and populations than it is to jump across content areas.

At the very least, the professional use of Radical Empathy in CDP work requires real mastery of the content being assessed – the standards and the KSAs that comprise each of them. Without that, it is just too difficult to imagine even a range of *successful* test takers.

Limitations of Radical Empathy

Radical Empathy is a rigorous and disciplined process and a useful tool. However critical it might be, it is no panacea. Even if a CDP could understand *every* possible cognitive path through an item, that would not necessarily tell them what they needed to do to improve the item.

Radical Empathy helps CDPs to recognize where there might be problems with items and which parts of the test taking population might run into them. However, it does not tell them whether those problems are severe or widespread enough to require remedy. That requires professional judgment.

Radical Empathy helps CDPs to spot problems with items, but does not necessarily identify solutions. It certainly does not select from the multiple potential solutions that might be most obvious.

Radical Empathy cannot even tell CDPs whether problems even have workable solutions.

Rather, Radical Empathy is an important tool. It should be applied early and often, and experienced CDPs should use it more or less automatically throughout item development. Because it is no magical panacea, waiting to think through the perspectives of *the range of typical test takers* can yield items that simply cannot be relied upon to support valid inferences about test takers, despite all the effort and resources invested into them.