

## Item Alignment Examination (Book Excerpt)

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.

# Item Alignment Examination

Item alignment is the building material for the foundation of all validity arguments for all tests, as we explain in Chapter [x]. Every plausible test purpose assumes that the test actually assesses what it purports to assess, and that requires the individual items to assess what *they* are purported to assess. If they do not, then the problems we discussed in Chapter [X] (i.e., construct underrepresentation, construct irrelevant variance, etc.) devastate test-content domain alignment and all test uses are based on false premises.

So, item alignment is pretty damn important.

Unfortunately, item alignment is not taken as seriously as it should be. First, this is because item alignment can *only* be considered by individually examining items for content. This cannot be automated, cannot be easily scaled, and requires expertise that does *not* generalize across content domains. Alignment examination requires content expertise. Alignment examination requires expertise in the cognition of test takers.

Item alignment examination requires the radical empathy with the perspectives of test takers of all item validity consideration. Not all test development professionals have that, nor do they need it. But CDPs often depend upon it, and nowhere more than when examining item alignment.

In this chapter, we present a methodology to perform item alignment examinations. Like all standard procedures and best practices, those new to it will feel that it is awkward and prescriptive. However, as CDPs gain proficiency with it and it becomes more habitual

through internalization, they will better understand its principles and they will become able to integrate item alignment examination into a dynamic process of item editing and improvement that keeps item alignment always present. Because item alignment is so important, until then, CDPs should make sure that they are as careful and deliberate in their examinations as the need to be to be sure that they are arriving at the most accurate understanding of items' alignment(s).

### **Overview of the Item Alignment Examination Procedure**

Item alignment is *not* found simply in the item, as it appears in print or on the screen. Rather, an item's alignment(s) are found by considering the task(s) that test takers embark on as they respond to an item. That is, the cognition, cognitive processes and cognitive steps that test taker engage in to respond to their understanding of an item. Validly aligned items elicit evidence of the *targeted* cognition. Examine items to determine that requires thoughtfulness, mindfulness and very rigorous thinking.

The basic process is as follows:

*0. Do Not Prejudge.* Do not look at the purported alignments, descriptions or rationales.

*I. Do It Yourself.* Work through the task prompted by the item as *you* would, were you encountering the item today, as who you are.

*II. Metacognition/Re-creation.* Reflect on *your* task, so that you can be mindful enough of it to avoid projecting onto others.

*III. Examine Your Atypicalness.* Look for how and why *your* task would be unlikely for actual test takers (e.g., test savviness, deep content expertise).

*IV. Radical Empathy.* Work through the item as a typical successful test taker's task and typical unsuccessful test takers' tasks

*V. Compare Tasks.* Revisit your task and the various tasks of your hypothetical test takers to identify the key KSAs that differentiate successful from unsuccessful responses to the item.

*VI. Interrogate Your Confidence.* Affirmatively look through all elements of the item for any suggestion that other KSAs and/or tasks might be relevant.

*VII. Confirm Alignments.* Look at the domain model or content domain definitions (e.g., content standards) for a highly aligned element(s).

*VIII. Consider Alignment Accuracy.* Compare the alignment(s) that you found with the purported alignment and the strength of the alignment(s).

This kind of close and careful examination of items is important at many stages of item development, but it is only for items that do not have obvious problems that need to be fixed. Similarly, it is not appropriate for initial intake of items because initial intake is not about final item validity. Rather, initial intake is about whether the item drafts have the *potential* to be turned into high quality valid items. Item Alignment Examination, on the other hand, is about increasing our awareness of the subtleties of how a fairly polished item may be received by test takers.

(Of course, this procedure is not practical when CDPs are not subject matter experts. This can happen with many professional certification and licensure exams, when the test assess very advanced or expert knowledge. Such CDPs should still understand this procedure, though they apply in a different way and in a different context. We discuss this in Chapter [X].)

### **Why *This* Procedure**

We have seen and experienced three common major problems with more casual attempts to identify item alignment. This procedure is designed to lower the likelihood of each.

#### **Projection**

Everyone projects their own feelings and thoughts onto other people. This is entirely natural. However, CDPs should be wary of projecting their own responses to items onto test takers.

First, CDPs are *quite* different than test takers. They are almost invariably far far more expert in the content domain and they have a level of test taking savviness that those with less experience with tests simply cannot approach. Thus, they can read items faster, spot key traits and elements of items more easily and immediately understand what items are asking for in ways that no test taker could.

Second, *every* individual – not just CDPS – brings their own experiences, views and identities to their interpretations and understanding of the world around them. One of us is a white man from the American Northeast and the other is white woman from the American South, and we can read things quite differently. But we recognize that our perspectives are usually quite similar. We are roughly the same age, with similar socio-economic background and overlapping cultural heritages. There is a huge array of potential areas, values, experiences and perspectives that could someone to understand and/or respond to a passage or item differently than one or both of us.

The first problem we see when people consider item alignment is that they project *their* understanding of the item and the task that *their* understanding prompts onto all potential test-takers. This is natural. This appears to be instinctive. And we like the idea that so many people have an immediate assumption that other people *are* like them. However, in this context, this is an enormous problem. As CDPs and other item reviewers are *not* typical test takers, this projection is clearly inappropriate. Furthermore, the more diverse the background and experiences of test-takers, the more open-minded CDPs should be about *multiple* different readings and responses to items by them.

### **Motivated Reasoning**

The second big problem we see is a common misunderstanding of what item alignment examination is about. At worst, it is treated as a pro forma exercise to check of a

requirement. While that is a rare extreme, all too often CDPs and reviewers fall short of the necessary skepticism about item alignment.

Most of the time, CDPs and item reviews are asked to *verify* item alignment, rather than assign it from scratch. In fact, because most items are written with particular targeted cognition in mind, it is quite rare for anyone to review an item that does not already have some recorded alignment(s). CDPs and other reviewers usually appear to assume that the recorded alignment is accurate, and only check it to make sure that it is plausible and not *obviously* off base. Sometimes, reviewers might not feel confident enough to challenge what is already recorded, but usually they fall short of giving deep and careful thought to whether the recording alignment is actually accurate.

Even when CDPs and reviewer *do* take a closer look, they assume that the item must be aligned with *something*. That is, if it is not aligned with this targeted cognition or standard, then there *must* be some other one on the list that the item *is* aligned with. This is a mistake. Item alignment is so important that it should never be assumed. The burden is on the item – or the CPDs and reviewers – to show that the item is, in fact, aligned with some appropriate targeted cognition. Item alignment is so important that every major round of reviews should build the case from scratch.

Failure to take this kind of care can let inappropriate items through. For example, items that were suitable for previous versions of the domain model (e.g., old versions of the standards) can slip through this way. More commonly, items that simply are easier to write, but never fit the requirements for *this* test can pass through because CDPS and

reviews do not realize that items can actual *fail* item alignment examination. That is, items that should be rejected because they fall outside the domain model are kept in the process, continuing to take resources away from more appropriate items.

### **Threshold**

Too many CDPs and item reviewers accept that an item is aligned with a particular targeted cognition if that targeted cognition is merely relevant to the item – part of any potential response to the item. But validly aligned items elicit *evidence* of the targeted cognition. As we explained in chapters [X] (*Theory of the Item*) and [Y] (*Validity and Alignment*), items only elicit such evidence when the tasks they prompt rely appropriately on the targeted cognition.

If there are alternative solutions to the item that do *not* rely on targeted cognition, then successful responses to the item do not tell us that the test taker has mastered the targeted cognition. If the targeted cognition is not the key and most challenging step of the prompted tasks, than unsuccessful responses to the item do not tell that the test taker has *not* mastered the targeted cognition. That is, if tasks do not *depend* on the targeted cognition as their *key* step, then performance on the item simply cannot provide evidence on the targeted cognition, and claim that such items are aligned with the targeted cognition are simply and inarguably false.

Frankly, it does a disservice to the hard work of domain modelers, educators and learners accept vaguely or tangentially relevant use of the targeted construct in some



potential tasks prompted by an item as sufficient for recognizing alignment. Doing so puts those parts of the content domain into a sort of second-class status, because they are not truly assessed, enough though the test *claims* to assess them. This can prevent the thoughtful work to find better ways to include those parts of the content domain on the tests.

### **Preempting Those Problem**

People naturally project, they naturally want to succeed and they are quite often willing to compromise. None of these are intrinsically bad. However, they can undermine item alignment and test-content domain alignment, and thereby undermine *all* further validity claims. The RTD Item Alignment Examination is designed preempt these kinds of natural mistakes and errors and to support thoughtful professionals in correctly identifying item alignment – or lack thereof.

This procedure is not designed for any particular Domain Model or set of standards, and it can be used when trying to simultaneously align items with multiple sets of standards (e.g., NGSS’s SEPs and IDGs). It simply requires carefully and consciously going through these steps.

### **RTD Item Alignment Examination Procedure**

#### **0. Do Not Prejudge**

Do not look at the purported alignments, descriptions or rationales. Do not look up the correct response, recorded difficulty or complexity, if you can at all avoid it. Do not

look at the meta-data, whether they are the supposed alignments or anything else. Do everything you can to approach the item as a test taker would, without forewarning of what is in it or what it might demand of you.

For many people, this zeroth step is the most difficult one. It requires a kind of discipline and intellectual integrity *not* to look at the answer when it is right there, available to be seen. It is hard not to satisfy your curiosity, and harder still not to rely on the recorded thinking of others to help you come to the best answer.

The most difficult form of this step is when the correct response is marked clearly in advance, and/or you already know the recorded alignments or other meta-data. This can occur because you have already worked with the item or because when someone presented to you they gave it away. “Hey, Marjorie, can you take a look at this? It says it’s a Standard 14b item, but I don’t see that. What do you think?” Even in this sort of case, you must put that knowledge out of your mind until the last steps of the this Item Alignment Examination procedure.

### **I. Do It Yourself**

Work through the task prompted by the item as you would, were you encountering the item today, as who you are. This should be the easiest step for most people, particularly if they were able to comply with Step 0. There is no need to put on another perspective, slow down your thinking, or do anything other than what you would do if you were faced with this item on a test.

So, respond to the item. Select or construct your answer, as your authentic self.

Check you answer to make sure it was correct.

If your answer was *not* correct, try to figure out where you stumbled, so that you can correct your mistake and work through a task to get to the correct answer. In this step, you do not have to think too much about how or why you went astray, as that is for Step II. Instead, just try to figure out how *you* might get to the right answer, before proceeding.

If you are not able to approach the item as a test taker then this step is just a little bit harder. Try to ignore what you know about the item and work through it as though you did not know it. In item development, the key (i.e., the correct answer option) is usually clearly marked, so you will need to ignore that. Keep in mind that that key may be misidentified, or the item might be doubled-keyed (i.e., have two correct answers), and this step is a good place to uncover that.

At this stage or the next, you may see such large problems with the item that it simply is not worth it to procedure with the rest of the Item Alignment Examination. If the item is clearly inappropriate and/or obviously has to seriously reworked, the subtleties that AIE is intended to reveal may be entirely changed by that work. One should only proceed with a full IAE if the item appears to be pretty good, as is.

## **II. Metacognition/Re-creation**

Having completed the item more or less naturally, go back and rethink what you did through that process. Examine your proficient, smooth and habitual thinking and

working through the task, slowing it down so you can identify every step and every KSA that you called upon. This can be challenging to do at first, but with practice it can become much easier. As you get accustomed to this kind of unpacking of and reflection upon your own thinking, you will inevitably start to look at items quite differently.

You should not merely try to recreate your productive path through the item and task. Rather, you should try to capture *every* step, be it productive, distracting or even just wrong.

- Were the instructions and stem immediately clear, or did you have to reread, revisit or slow down to understand some aspect of them? When you work through your task
- Did you start over at any point, or even move back a few steps?
- When, if at all, did you glance back at the instructions or the stem?
- Did you need double- or triple-check anything at any time.
- If there was stimulus, how did you move back and forth between the stimulus and the rest of the item as you worked through it?
- Did you read or check ahead, or double back, at any time?
- How did you work your way through the selected response options? Did you read them all before selecting one? Did you jump to one of them quickly? Did you systematically work your way through all of them? Were some require a little more of you to understand them?

- Were some of the distractors (i.e., incorrect answer options) easier or harder to discard than the others? What made them different from the others?
- Did you have an answer in mind before looking at the answer options, or did you use the answer options to help you to come to some sort of answer?
- What was the cognitive path that *you* took to respond to the item, however, circuitous? That is, what was the task that *you* built to respond to the item
- What KSAs did you use – even if you discarded them – through this task? Which could we easily take for granted in test takers, and which might not present any level of difficulty for some test takers?
- Focusing on those key steps and KSAs that might present barriers to some test takers, think about their complexity, difficulty and obscurity for you – or for remembered you.

This step is absolutely critical to understanding the perspective of test takers and simply cannot be skipped or rushed. If you do not consciously recognize and own your own process, it will be nearly impossible to avoid projecting elements of it onto others. You must reflect on *your* task, so that you can be mindful enough of it to avoid unconsciously projecting it onto others.

While addressing the projection problem would be reason enough for this step, it is not the only reason. This kind of deliberate examination of what it can take to respond to the item and *why* you took the path you took acts as a sort of trail breaking for your later

thinking (i.e., Step IV) about how others might respond to the item. As you become more conscious of the little recognitions and decisions that made up your thinking, you build a model that can make more visible to you places where others might think differently – either productively or unproductively for them. Becoming more aware of the complexity of your own thinking serves to increase your ability to recognize or imagine the thinking of others.

### **III. Examine Your Atypicalness**

Having uncovered and re-created your thinking through the item, you must now interrogate it and recognize the ways in which is an unlikely -- or perhaps even implausible – response to the item by authentic test takers. You are not a typical test taker, and you must recognize how that makes your response to the item an exceptional one.

CDPs and item reviewers are quite different than expected test takers. Certainly, CPDs should be very savvy test takers, with unimpeachable expertise in how items work and what they might be looking for that. That is, they never suffer for unfamiliarity with test or item formats. CPDs and item reviewers usually have long experience with high levels of content expertise. This does not simple make them like high performing test takers, as for them, they have developed sufficient proficiency that many KSAs have become habitual skills and well-integrated knowledge. This lowers cognitively complexity and speeds up discarding of some distractors. Furthermore, CDPs knowledge of the domain model – and knowledge of which parts of it are being targeted by items – can

prime them towards particular tasks that authentic test takers might not recognize as quickly, if at all.

Of course, test takers respond to items in situations of *at least* somewhat elevated stress. This can make them more likely to rush or misread what is in front of them in ways that CDPs and item reviews are far less likely to.

In addition to all of this, you might have some special knowledge or personal experience – good or bad! – that makes you respond to some items a big differently than others, even other CDPs.

Consider all of this and look back on what you uncovered in Step II. Try to be mindful of how your atypicalness may have influenced how you read the item, how you built your task and how your worked through it. Make doubly sure that you understand how your response to the item was rather unlike what one might expect from authentic test takers.

This step can be quite

#### **IV. Radical Empathy**

Work through the item and a typical successful test taker's task and typical unsuccessful test takers' tasks is the central Step of RTD Item Alignment Examination. It requires putting on the thinking persona of test takers – who are necessarily quite different than yourself. It requires sensitive to the diversity and range of test takers, both in backgrounds and experiences, and in potential responses to item. Accurately recognizing

item alignment(s) means recognizing the key steps that lead some test takers to respond successfully to items and the key steps that lead other test takers *not* to respond successfully. All of that calls a kind of radical empathy with test takers that is critical for content development work.

For this step, essentially repeat what you did in Step II (*Metacognition/Recreation*) – carefully and deliberately working through the item and task – but from assumed perspectives, rather than your own.

The easiest test taker perspective to take on is someone much like yourself at that age, or were you in that position – a successful test taker, of course! Rather than taking your expect path through it, use what you learned in Step III (*Examine Your Atypicalness*) to construct a response to the item as a successful test taker. Be as conscious and as careful in thinking through this as you were when recreating your own path. Again, note the key steps of this response to the item, where this hypothetical test taker drew on the more challenging or newer KSAs or cognition.

Then, work through item again, putting on still other perspectives. How much less successful test takers work through the item. Intentionally build in the kinds of mistakes that you think these test takers might make – from being thrown off by some element of the item to plausible misunderstanding of the content domain. What might happen were a student to rush? Carefully think about other paths that a test taker might take, in response to the item. Take note of the mistakes you add, how the hypothetical test takers work from



there, and whether the available answer options indicate that s/he should go back and start again.

Hopefully, by having gone through multiple rounds of these assumed perspectives you have worked to each answer option – the key and all of the distractors. However, if not, try to figure out how a test taker might arrive any that remain.

The hard part of this step is truly putting on different personas, trying to assume a radically different identity, background and/or experiences and to capture how a test taker so different than you might respond differently to the item. This gets to fairness issues of bias and sensitivity, of course (see chapter [X]). Fairness is simply too important to put off until later stages of item development, and it should be considered during the earliest rounds of item review by CDPs. Knowledge and sensitivity to the targeted testing populations its various subgroups is invaluable here. You should also affirmatively look for elements of the item that might strike some test takers differently than others.

This step is truly one of the most important tools or skills in any CDP's toolbox. Being able to see items through the eyes of test takers – through the eyes of many different types of test takers – is simply vital understanding the items as a content development professional. This is the only to understand what items are assessing, to spot and to correct deep problems in items, and to ensure that you are producing items that elicit evidence of the targeted cognition.

No, this kind of radical empathy with a wide range of test takers it is not easy – certainly not at first. It is a skill that needs to be developed, starting with types of test

takers that you have more in common with and quickly and intentionally expanding to a broad array of other kinds of test takers. Over time, you will become better and faster at it, and even able to think about multiple perspectives at a time.

## V. Compare Tasks

Having built up your lists – however formally or informally – of the KSAs and cognition that you think test takers might attempt as they respond to items, it is *finally* time to think about the actual question of alignment. Do not forget, however, that items are not aligned with a KSA or targeted cognition simply because some test takers might make use it. Rather, *valid items elicit evidence of the targeted cognition*, which means that test taker performance on an item should provide specific affirmative or negative evidence of their mastery of the targeted cognition. This requires the aligned content to be the *differentiating factor* between those who successfully respond to the item and those who do not.

Therefore, the knowledge, skills and abilities that test takers tap into, regardless of their success or failure with the item are relevant, and are what ECD calls *additional KSAs*. However, none of those are the item alignment.

Knowledge, skills and abilities that some successful test takers tap into and that some unsuccessful test takers tap into are *also* merely *additional KSA*.

The actual item alignments are the KSAs and/or cognition that successful test takers make use of and unsuccessful test takers do not or were not able to make use of.

This can include strategic thinking in task selection. This can include specific knowledge or skills that successful test takers apply correctly and unsuccessful test takers apply incorrectly. This can include specific misunderstandings that lead unsuccessful test takers astray and do not impact successful test takers. It can even include fairness, bias and/or sensitivity distractions that are far more likely to arise for some test takers than others (see chapter [x]).

Items of Aristotelian perfection would give you just one perfect KSA, one piece out of the domain model, one content standard. It would be simple and clear exactly what differentiates successful test takers from unsuccessful test takers. And some items really are able to reach this standard, particularly those aimed at raw application of simple skills or very particular knowledge. However, most items are not that perfect. Between the greater complexity of KSAs and/or particular nature of the item in question, there can be a longer list of differentiating factors.

## **VI. Interrogate Your Confidence**

Step VI is about humility and double checking your work by examining your earlier results from an different angle. Having already thought deeply and deliberately about the experience of working through the item from multiple perspectives, step back and examine the item itself. Having already tried to think like a test taker and work through tasks towards an answer, put your editor or analyst hat back on to break down elements of the item *directly* for the purpose of flagging those that suggest particular KSAs or cognition.

Affirmatively look through all elements of the item for any suggestion that other KSAs and/or tasks might be relevant.

Look for key words, phrases or symbols in the instruction and/or stem that point to the key/differentiating KSAs and/or cognition. Some can be immediately obvious, while others might be a little bit more subtle or vague while still directing the test taker. Similarly, look for key words or phrases that suggest *other* KSAs and/or cognition. What do those parts of the item seem to suggest?

Do the same for the key. Are there ways in which the key somehow stands out from the distractors? Does it somehow suggest or call out for any particular approach? Do the distractors suggest an approach, with all of its KSAs and/or cognition? If the item has a stimulus, examine the stimulus for elements or aspects that might suggest particular KSAs and/or cognition, too.

In this step, you are trying to be savviest possible test taker, using the item to try to read the intentions of item's developers. Without actually attempting to solve or respond to the item, try to read the item for what it *looks like* item developers are trying to get you to do. Of course, this is a bit more difficult when you are one of those developers, but it is no more difficult than trying to read one's own writing for revision in other forms.

The work of this Step is a vastly inferior approach to identifying item alignment, but unfortunately many have used it as their dominant – and even only – procedure, in the past. It simply is no substitute for actually working through the item and the task it

prompts as a test taker. It is no replacement for the radical empathy of Step IV. However, because it is such an different approach, it is a good check on your earlier work.

Compare what you have seen in this review of item to your list of key differentiating KSAs and/or cognition.

- Does the language and presentation of the item suggest particular KSAs and/or cognition that are not on your list? Did you discard them because your analysis found them *not* to be differentiating of successful and unsuccessful test takers?
- Does it suggest particular KSAs and/or cognition that you never tapped into in any of your attempts to work through the item as a test taker? Knowing this, can you see what that or those tasks and figure out whether they might be differentiators?

Most of the time, you will not uncover new KSAs and/or cognition this way, certainly not as you get better at the radical empathy of Step IV. However, this is like proofreading your work, one last time. Even if you do not think you will find any mistakes, it important to make sure.

## **VII. Confirm Alignments**

Finally, you get to actually consider the labeled alignment(s) of the item. Now that you have so carefully done your own work to identify the differentiating KSAs and/or cognition, you are in a position to assign your own labeled alignment(s). Of course, this depends on familiarity with the domain model, content domain definitions (e.g., content standards) and/or assessment targets, as that is what labeled alignment(s) must reference.

Do *not* assume that that the key differentiators will actually appear in the content model. As discussed above, this is a common mistake and undermines test-content domain alignment, severely. We simply cannot caution you strongly enough, you must find an actually good fit between your key differentiators and the elements of the domain model. If you know the domain model well, you likely already have expressed the KSAs and/or targeted cognition in the language of the domain model, making this a bit easier. If you are still learning the domain model, be careful not to be subject to the kind of motivated that has you understanding its elements based on what you are looking for, instead of what is actually expressed.

Work through your list of key differentiators and put them in the language of the domain model, where you can in good faith. Those are the item's alignments. Again, ideally you will have just one alignment, in many cases. For tests whose items each are supposed to be aligned with two different kinds of standards, ideally you will find just one from each list. However, you will often find more than one.

Therefore, do not stop looking for alignments when you have found the first one. Keep in mind that your key differentiators might, in fact, reasonable be seen as multiple different standards or elements of the domain model. Perhaps it is Standard 11b.5, with list a little bit of this extra KSA in one view, but Standard 5a.2 with a bit if this KSA in another view. If you found multiple truly different tasks to arrive at answer, you might simply have multiple different alignments, depending on how the test takers respond to the item.

Hopefully, you will find that that the previously labeled alignment is accurate. However, make sure that it is not based on a misinterpretation of the standard or any other kind of shoehorning. And make sure that you have shoehorned your key differentiators into any potential alignments. Just because you think that your key differentiators are important to the content domain does not mean that it actually appears in your domain model or assessment targets. For example, in recent years, ELA standards have *not* included the ability to recognize the part of speech of a word in a sentence (e.g., Is it a noun? Is it an adverb?). This has long been a staple of many English classes, and therefore both English teachers and those who grew up with those lessons are sure that there *must* be a standard for such items. They end up shoehorning these items into inappropriate standards because they are unwilling to let go of something that has long been important to them, even though it is not their job to rewrite the standards.

### **VIII. Consider Alignment Accuracy**

The last Step of the RTD Item Alignment Examination depends deeply on professional judgment. There are no hard and fast rules, and no clear thresholds. Rather, you must consider the expectations of your clients, the requirements of your project and the purpose(s) of your test. How strong should a labeled alignment be to pass muster? How complete? How much of a problem does it present if there are differentiators that are not included?

It is even more difficult to figure out what to do when you have found that an item can prompt multiple quite distinct tasks to get to the answer option. If one of them would only be attempted by a tiny tiny tiny minority the test population, it is not hard to disregard it, but the basis for such judgments is difficult to be sure of. This leaves the question of how prevalent some key differentiators might be – as compared to others – on the table.

Thus, you must depend on professional judgment, collaboration and understanding various needs and constraints for your project when ultimately deciding how to label many items' alignment(s). The practical work of test development – like most all practical work – eventually relies on compromise between different values and goals and on a sense of what is good enough for this piece of this project at this time. That is not license for simply abandoning all sense of accuracy and/or validity, of course. Incorrectly labeled item alignments undermine the purpose and use of *any* test. Rather, it simply means that you cannot let the perfect be the enemy of the good and you must pick your battles so that you do not lose your credibility and/or seat at the table over minor issues of relatively small importance.

### **Using the RTD Item Alignment Examination Procedure**

As we have said above, this process might feel cumbersome at first. It certainly is more difficult and more work than simply jumping to Step VI and relying on shoehorning. However, it has the benefit of actually attempting to uncover what the items *actually*



assess, instead of simple what we might want them to assess. As you become more adept with this approach, it becomes must easier, more natural, and even gets integrated into your thinking every time you look at an item. You will gain easier access to the kind of radical empathy that is entirely necessary for producing high quality valid items and high quality tests.

