

Chapter 2

A Framework for Improving the Accessibility of Assessment Tasks



Erik Roelofs

Abstract In constructing tests it is vital that sources of construct irrelevant variance are minimized, in order to enable valid interpretations about the test taker. One important source of construct irrelevance is inaccessibility to the test and its items. In this chapter a framework is presented for design and review of test items, or more broadly, assessments tasks, to ensure their accessibility. An application is presented in the context of theory exams to obtain a drivers' license in the Netherlands.

2.1 Accessibility of Assessments

Access in the context of educational testing refers to the opportunity for a student to demonstrate proficiency on a target skill (e.g., reading, mathematics, science). Accessibility is not seen as a static test property but instead, as the result of an interaction among test features and person characteristics that either permit or inhibit student responses to the targeted measurement content (Kettler et al. 2009; Ketterlin-Geller 2008; Winter et al. 2006).

In more general terms, accessibility is considered as a prerequisite to validity, the degree to which a test score interpretation is justifiable for a particular purpose and supported by evidence and theory (AERA, APA and NCME 2014; Kane 2004).

Limited accessibility can have various causes, that may threaten the validity score interpretation in different ways, depending on the assessment purpose.

A first source of limited accessibility pertains to the situation where test takers do not yet master the target skills. If this is the case, a conclusion might be that the test taker needs to go through an extended or adapted learning process. There would not be a threat of validity. However, if the test in general was judged as too difficult for the intended target group, this could be a case of misalignment between test content and intended outcomes, which can be considered as a threat to validity.

E. Roelofs (✉)
Cito, Arnhem, The Netherlands
e-mail: Erik.Roelofs@cito.nl

Second, a lack of ‘hard’ access capabilities is a well-documented source of access problems (Sireci et al. 2005). Due to some sort of disorder or handicap a student is not able to process task information or to respond to the task. For instance, test takers with a visual or auditory impairment, a motor disorder, ADD, autistic spectrum disorder, or dyslexia, may lack necessary capabilities to have access to a test item. For these test takers, the validity of inferences is compromised when they cannot access the items and tasks administered to them and with which they are required to interact. To improve accessibility for students with special needs, an example of reducing barriers and increasing accessibility, is the provision of a read-aloud assessment administration.

A third source of limited accessibility is lack of access skills that can be developed through education, but of which it is questionable whether they belong to the target skill or competency (Abedi and Lord 2001; Kettler et al. 2011). For instance, assessment tasks for math may place a high burden on reading skill and may subsequently cause access problems. The question is whether students in this case do not have the necessary level of target skill or whether they lack the access skill of reading comprehension. Some of the accessibility problems can be prevented by defining the target skill related to the assessment purpose well in advance, and determine whether access support is necessary, without impacting the measurement of the target skill.

A fourth source of limited accessibility is related to flaws in task presentation itself, that may either result from constructors’ mistakes or by misconceptions about a task presentation feature. By their very design assessment tasks themselves can be inaccessible to some and even all students. Design flaws relate to errors, inconsistencies, omissions in the assignment or in the task information, or in response options and cause extra processing load for students (Beddow et al. 2008).

In the remainder of this chapter we concentrate on the design principles for improving assessment accessibility, that help to avoid unnecessary, construct irrelevant task load, and thereby improve the validity of claims about test takers’ target skills.

2.2 Principles that Underlie Accessible Assessment Design

2.2.1 Principles from Universal Design

To address the challenge of designing and delivering assessments that are accessible to and accurate for a wide range of students, principles of universal design (UD; Mace 1997) have been applied to the design, construction, and delivery of tests. The core tenet of UD is to create flexible solutions that avoid post hoc adaptation by considering from the start the diverse ways in which individuals will interact with the assessment process. Dolan and Hall (2001, 2007) therefore proposed that tests be designed to minimize potential sources of construct-irrelevant variance by supporting the ways that diverse students interact with the assessment process. Thompson et al. (2002) adapted Mace’s original elements from architectural design to derive seven elements

of accessible and fair tests: (1) inclusive assessment population; (2) precisely defined constructs; (3) accessible, nonbiased items; (4) items amenable to accommodations; (5) simple, clear, and intuitive instructions and procedures; (6) maximum readability and comprehensibility; and (7) maximum legibility. Ketterlin-Geller (2005, p. 5) provides a more generic definition of universal design for testing: an “integrated system with a broad spectrum of possible supports” that permits inclusive, fair, and accurate testing of diverse students.

2.2.2 Principles from Cognitive Load Theory

In cognitive load theory it is assumed that any instructional task, including an assessment task, has a certain amount of intrinsic task load. This is the natural complexity that the task possesses, due to the complexity of the task content (Sweller 2010). In Fig. 2.1 this is depicted by the green part of the left stacked bar, representing the total task load. The intrinsic task load is strongly related to the number of task elements that need to be processed simultaneously by a test taker. However, test items may also address learnable access skills (yellow colored area) which are not part of the target skill per se, such as reading a text or understanding a picture. In addition, the test item may address hard capabilities (red shaded area), such as eyesight, quality of hearing and short term memory. Depending on the target group, these types of task load may be seen as inevitable, or as avoidable or even irrelevant. The part of the load that can be avoided is hypothetically depicted through dotted areas of the left bar. A final source of extrinsic task load includes load that is caused by errors or flaws in parts of the item itself (blue dotted area). This source of extrinsic task load should always be avoided.

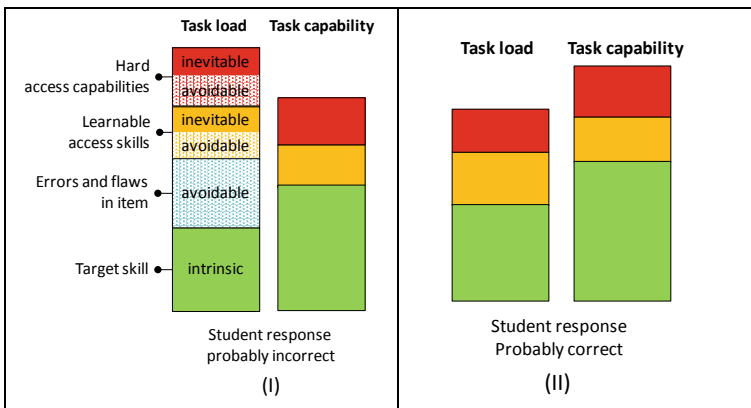


Fig. 2.1 Intrinsic and extrinsic task load versus test taker task capability in an assessment task

A test taker solving an assessment task is confronted with the total task load of an item and uses his or her own available combination of hard access capabilities, access skills, and target skills, his or her ‘task capability’. Part of the task capability level is the ability to deal with avoidable extrinsic task load caused by the assessment task presentation over and above the level of the intrinsic task load, which is handled by applying the target skill. However, in the hypothetical situation in Fig. 2.1 (part I) the test taker will probably not be able to succeed because his task capability level falls short of the total level of task load.

When, however, all avoidable sources of extrinsic task load have been stripped off, and the amount of intrinsic task load remains constant, the test taker’s level of task capability in this hypothetical situation is high enough (see part II of Fig. 2.1), to solve the task correctly.

Extrinsic task load in assessment tasks (test items, assignments) is for a large part caused by the way a task is presented. As discussed, test takers are forced to use skills that are not necessarily related to the target skill, but which do require cognitive resources. According to cognitive load theory (CLT) different cognitive mechanisms can be involved in causing extrinsic load.

- Quantitative overload: more information is presented than is strictly necessary for the assessment task.
- Deception: information sets off thinking directions that prevent test takers from arriving at a (correct) solution.
- Missing information: the test taker cannot get a complete problem representation, due to missing information.
- Split attention: test takers have to keep elements of information from different locations in their short term memory, to be able to respond to the task.
- Redundancy effects: the task presents redundant information that causes an extra memory load. E.g. redundant verbal explanation of an already self-evident picture.
- Interference-effects: verbal and visual information contradict or differ in nature, causing additional processing load for the test taker.

In the next section we present a framework for item design and item screening in which we intend to separate sources of extrinsic task load from intrinsic task load. Moreover, guidelines are provided to reduce extrinsic task load in different parts of an assessment task, taking the perspective of the test taker who solves a task.

2.3 Evaluating and Improving Accessibility of Assessment Tasks from a Test Takers’ Perspective

In order to prevent unnecessary extrinsic task load it is important to evaluate assessment tasks (items, assignments) on possible sources of inaccessibility as described above, and modify these tasks accordingly. For that purpose two excellent tools have already been developed, the Test Accessibility and Modification Inventory (TAMI; Beddow et al. 2008) and the TAMI Accessibility Rating Matrix (ARM; Beddow et al.

2013). Beddow et al. (2011) report about how the use of the tools in the context of biology items improved the accessibility, while at the same time the psychometric quality of the items was improved.

The TAMI-tool used to evaluate and improve item accessibility is predominantly structured into of five categories, that represent the typical format of items: (1) item stimulus, (2) item stem, (3) visuals, (4) answer choices, (5) page/item layout. In addition, TAMI has a sixth dimension, referred to as Fairness, which refers to a broad collection of issues, such as respect for different groups, the risk of construct-irrelevant knowledge, emotion or controversy arousing content, the use of stereotypes, and overgeneralizations.

We extended the TAMI-model to use it in the context of item development and item piloting in two ways (Roelofs 2016). First, the structure of the screening device is built around the cognitive and self-regulative processes that test takers engage in, when solving an assessment task. By doing so, the item writer takes the perspective of the test taker. This idea stems from the cognitive labs methodology, which is used during item piloting. The methodology employs procedures, such as thinking aloud protocols by test takers, that provide insight into respondents' thought processes as they respond to test items. In cognitive laboratory interviews evidence is gathered to ensure that target audiences comprehend task materials—survey scenarios, items/questions, and options as they were designed. If the evidence suggests that task materials are not being understood as survey and test developers designed them, then modifications can be made in the wording of questions and options to ensure that the correct interpretation is optimized (Leighton 2017; Winter et al. 2006; Zucker et al. 2004).

Second, in addressing the accessibility of parts of an item, we deliberately separated guidelines in order to create intrinsic task load and to prevent extrinsic task load. This is done, because the question what content belongs to the target skill, intrinsic task load, and what belongs to the access skill is often a matter of definition. In solving mathematical word problems, for instance, reading can be seen as part of the target skill, as an inevitable necessary access skill, or as a skill that may hinder task solution, if not mastered at a proper level (Abedi and Lord 2001; Cummins et al. 1988).

An accessible assessment task supports the test taker to perform at his best, without giving away the answers on the items. Our heuristic model (see Fig. 2.2) depicts the test takers' response process to assessment tasks in a series of five connected task processes. The model is similar to the Winter et al. cognitive lab model (2006), but was adapted to account for the self-regulation activities, including monitoring, adjustment and motivation activities that take place during solving tasks (Butler and Winne 1995; Winne 2001). The processes in our model of task solution entail: (1) orientation on the task purpose and requirements, (2) comprehending task information, (3) devising a solution, (4) articulate a solution. During these task processes, that need not be carried out strictly sequentially, a self-regulation process including planning, monitoring and adjustment takes place (Winne 2001). The model of test takers' actions and thoughts is heuristic and general in nature and can be applied to different subject domains.

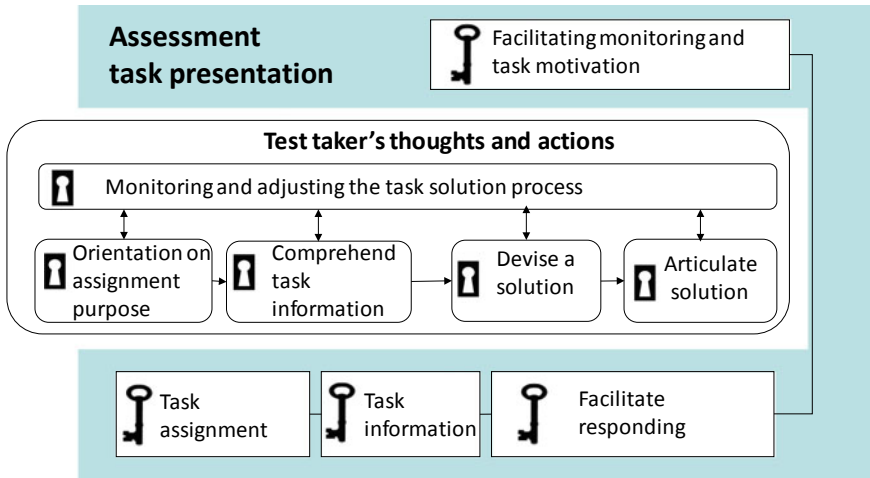


Fig. 2.2 How assessment task presentation acts on test takers' thoughts and action the during task solution process

From an accessibility perspective it is the item constructors' task to support access to all sub processes of task solution by means of an optimal assessment task presentation. This involves the second ordering structure in our model and relates to similar categories as used in the TAMI-model, although reworded to cover any type of assessment task: assignment, task information, response facilitation, and lay-out/navigation.

Using this model and building on the recommendations by Beddow et al. (2008) we developed a checklist of requirements by which the most common problems of task presentation that act on test takers' response processes can be addressed (Roelofs 2016). The content of the recommendations and the checklist are summarized in Tables 2.1, 2.2, 2.3 and 2.4.

2.3.1 Supporting Orientation by a Clear Assignment

During orientation on an assessment task the test taker tries to find out the required outcomes of the task and expectations about the way these outcomes can be achieved. A clear assignment will support this orientation, including a clear statement about the product, the approach, the tools to be used, the criteria of quality and accuracy to be taken into account.

In order to create appropriate intrinsic task load in the presentation of the assignment it is essential that test takers are stimulated to show intended target behavior, including the steps to be taken, strategies to be used, tools to be used, outcomes to be accomplished, according to a specified level of accuracy. Reducing extrinsic load

Table 2.1 Guidelines for presenting the assignment

A1 Assignment: Provide information about the expected product, activities to be carried out, the purpose of the task

Create intrinsic task load

- Make sure the assignment is logically aligned with the task information
- Make sure the assignment fits with the aimed target behavior, needed to demonstrate the skill

Reduce extrinsic task load

- Use specific words. Limit the number types of information
 - Length: keep assignment as concise as possible
 - Align the thinking direction of the task information with the assignment
-

A2 Guidelines for task execution: provide guidelines for a requested line of action, thinking steps to be shown, refer to the use of response tools and resources (see R4)

Create intrinsic task load

- Make sure that guidelines stimulate test takers to show intended specific target behavior, such as steps, tasks, strategy use, outcome, product
- Clarify which tools and resources can or must be used (e.g. scrap paper)

Reduce extrinsic task load

- Give clear directions about the approach to be used. Make sure the requested responding behavior comes naturally, by facilitating response behavior instead of long explanations
-

A3 Task outcome requirements: clarify requirements for presentation, quantitative and qualitative outcome requirements, accuracy of solutions

Intrinsic task load

- Align requirements with intended task authenticity and performance conditions
- Align accuracy of solutions or the result (number of words, accuracy margins) with what is intended

Reduce extrinsic task load

- Provide clear and concise directions about performance requirements
-

in orientation is accomplished by optimizing the presentation of the assignment. It is important to align the thinking direction of the assignment with the task information, the use of clear, concrete and concise language in giving directions. In general, we content that it is better to have self-evident response methods than to use (long) explanations of how to respond.

2.3.2 *Supporting Information Processing and Devising Solutions*

During the process of comprehending task information (second process) and the process of considering and devising solutions (or answers, third process), the test taker actually processes information that is relevant for solving the task, including some starting situation. Decisions are made about which information is essential for task solution, and strategies are chosen and applied to arrive at a solution, either in separate steps or by directly responding. Task information to be presented can relate to multiple objects, persons, symbols, actions that the test taker needs to act on. These can be presented within different contexts, that differ in authenticity, and conveyed

Table 2.2 Guidelines for presenting task information

I1 Task information: Present information from/about objects, persons, situations, that the test taker needs to act on. Align complexity level with the target group level

Create intrinsic task load

- Task information enables the onset of the target skill
- All necessary task information is given to carry out a task
- The amount of task information is balanced with the size of the task
- Level of complexity of task information is aligned with the developmental level of the target group

Reduce extrinsic task load

- Prevent mental overload, misguidance or deception, including:
 - Too high text complexity for the task purpose or the target group
 - Redundant textual information
 - Misleading textual information
 - Unnecessary complex visual information
 - Irrelevant visual information
 - Conflicting information
-

I2 Contexts: use contexts that are essential to the target skill and the target group

Create intrinsic task load

- Presented actors, objects, materials apply to the intended task situations
- Task contexts and their levels of authenticity align with what is intended
- Use context only when it is functional for demonstrating the target skill
- Choose the necessary level of context authenticity: real, simulated, no context

Reduce extrinsic task load

- Prevent the use of distracting, non-functional context-information
 - Prevent confusing, stereotyping, exclusive, implausible contexts
-

I3 Media/stimuli: choose media that convey necessary task information faithfully

Create intrinsic task load

Make sure the media and stimuli represent the targeted task situation needed to demonstrate a target skill. E.g. the intended reading text types, math contexts, their presentation (plain text, web page, video clips, drawings, photos, animations)

Reduce extrinsic task load

- Provide maximum perceptibility of images, sound, speech
- Provide readability of characters, printed text, font
- Usability of media: adapt controls until these work intuitively
- Availability of tools to perceive stimuli (e.g. microscope, headphones)

Enable rewinding audio-visual stimuli, unless target skills require only one attempt

I4 Problem definition and support aligned with the target skill and the target group

Create intrinsic task load

- Define and structure the problem in line with the intended target skill, the target group, and desired performance conditions
- Determine how much regulative support is given to a test taker, based on intended level of independent task performance. Determine about supporting information, scaffolding of solution steps, or hints

Reduce extrinsic task load

- Present support and hints concisely and prevent spread-out of information
-

Table 2.3 Guidelines for facilitating responding**R1 Target material that the test taker has to act on***Create intrinsic task load*

- Choose the material to be acted on: information, objects, materials, devices, persons, to the intended target behavior, task situations, and the level of authenticity

Reduce extrinsic task load

- Delete unnecessary target material, which is not to be acted upon in a task

R2 Interactivity of responses*Create intrinsic task load*

- Determine whether the onset of the target skill requires an interactive response with visible consequences for: the actor, recipients, objects, present in the task context. Examples: marking of sentences, moving an object, an action with a visible result

Reduce extrinsic task load

- Make sure no unnecessary interactions can be carried out during responding

R3 Support in responding*Create intrinsic task load*

- Make sure the level of support in responding corresponds with the intended performance conditions, as regards the target skill and the target group

Reduce extrinsic task load

- Prevent that support is redundant or oppositely too concise or even distractive

R4 Strategic and technical tools*Create intrinsic task load*

- Add all necessary strategic and technical tools that enable test takers to express the target skill, in line with requirements about: the features of task situations; authenticity of the task context; specified performance conditions

Reduce extrinsic task load

- Make sure the tools work as expected and do not require highly developed handling skills

R5 Response mode and sensory channel*Create intrinsic task load*

- Use a response format that enables the best expression of the target task behavior
- Use the most appropriate response mode that expresses relevant task behavior (verbal, visual, psychomotor)
- MC-options contain necessary information, are plausible, congruent with stem and task information, logically ordered, distinguishable; contain one correct option

Reduce extrinsic task load

- Replace indirect response modes by direct modes
- Prevent the necessity to make mode switches or direction changes in responding: from verbal to visual; non-corresponding orders
- MC-items: construct distractors that do not cause confusion or mental overload

R6 Response tools*Create intrinsic task load*

- Choose response tools that enable a natural expression of a target skill. The required behavior in responding to tasks is supported by tools like drawing tools, formula-editors, word processor, chat apps, game controls

Reduce extrinsic task load

- Make sure the test taker can express the answer fluently and intuitively
- Prevent any technical obstacles or difficulties in operability of response tools

Table 2.4 Guidelines regarding item lay-out and test navigation**SR1 Perceptibility-readability of screens, pages, item parts***Create intrinsic task load*

- Presentation is aligned with the intended situation in which the test taker is supposed to perform
- Consider: to what extent does increasing perceptibility of task information change the necessary level of intrinsic task load for measuring the target skill

Reduce extrinsic task load

- Make sure that screens, pages, item parts have a recognizable lay-out and content
- Make sure the test taker does not get stuck in the task, because of the chosen level of perceptibility, readability of item parts, causing mental overload or confusion

SR2 Lay-out of item parts*Create intrinsic task load*

- Make sure that presentation of the task elements represent task features of the target situation for the intended target group
- Make sure that improving the accessibility of information by structuring its presentation does not change the necessary intrinsic task load that is part of the target skill (e.g. in reading comprehension)

Reduce extrinsic task load

- Make sure that the test taker does not get stuck in the task, because of the chosen lay-out and structuring of task elements, causing mental overload (e.g. spread of information, confusing tables, schemes)

SR3 Navigation through the test*Create intrinsic task load*

- Make sure that the type of navigation support provided to the test taker represents the target situation for the intended target group

Reduce extrinsic task load

- Make sure the test taker does not get stuck in the task, because of a lack of support in monitoring own progress in item solving and navigation

through different modes of presentation, that may address different senses resulting in different processing.

In order to create appropriate intrinsic task load (see Table 2.2), it is essential that information is necessary for the onset of the target skill, including solution processes at a level that fits to the developmental level of the target group. Media need to deliver a faithful presentation of the information. The degree of problem definition and structuring is need to be aligned with what is intended for the target group.

Reducing extrinsic task load in task information includes avoiding the use of distracting, non-functional context-information, including the use of confusing, stereotyping, exclusive or implausible contexts. Images, sound, and speech should be maximally noticeable. Printed characters, fonts need to be well readable, media controls should work intuitively, and if necessary, tools to perceive stimuli (e.g. microscope, headphones) need to function well.

2.3.3 Facilitating Responding

In the fourth phase the test taker prepares and expresses responses on the assignment. It needs to be clear on which target material he or she needs to act, how and in which mode is to be responded, and which sensory channels should be used, with or without the use of strategical tools (e.g. dictionaries, search engines), and technical tools (E.g. calculators, rulers, carpentry tools).

In order to create appropriate intrinsic task load (see Table 2.3) it is essential that all required facilities for responding activate the behavior necessary to observe the target skill in the intended task situations. This relates to the choice of material to be acted on, the level of support given (e.g. hints, structure), the format and the sensory channel that best expresses the intended behavior. Response tools should enable a natural expression of that behavior.

Extrinsic task load is reduced by stripping all sources of distraction, confusion, or mental overload. This can be done by deleting material, that is not to be acted upon in a task, avoiding unnecessary interactions, avoiding response channel-switches (from verbal-to visual), avoiding non-corresponding orders between task information text and (multiple choice) response options, using direct response modes. Finally, the response tools themselves should function technically fluently and intuitively.

2.3.4 Facilitating Monitoring and Adjusting

During all phases of responding, an overarching process of monitoring and adjusting of the task solution process takes place. In various domains, these self-regulation skills are seen as an inseparable part of the target skill, e.g. reflecting and revision in writing (Deane 2011).

In general, the presentation of the assignment already implies support of self-regulation: orientation on task requirements and anticipated task outcomes form the basis of monitoring by the test-taker. In addition, lay-out of the individual assessments tasks and the navigation through the test can support self-regulation on the part of the test taker.

In order to create appropriate intrinsic task load it is essential that navigation design and item formatting optimally support information processing, responding and self-regulation. This is to be achieved through presenting information that is aligned with the intended situation in which the test taker is supposed to perform.

In order to prevent extrinsic task load it is essential that the test taker does not get stuck in the task, because of low perceptibility, readability of item parts, causing mental overload, or confusion. A well-known phenomenon is that of split attention, which takes place when test takers have to look for information elements that are spread over different locations in order to respond to the task (Ginns 2006). Finally, navigation can be assisted by providing feed-back about the progress status (status

bars, finished and unfinished items) and by providing tools to navigate through the test without getting lost.

2.4 An Application of the Test Accessibility Framework: The Dutch Driving Theory Exam

In this section we describe an application of our accessibility improvement framework in the context of the Dutch driving theory exam.

2.4.1 Innovations in the Dutch Traffic Theory Exam for Car Drivers

During the last decade the Dutch theory exam for prospective car drivers (B license) has undergone major changes. First of all, driving a car is seen as a complex task that is inter-twined with daily life tasks. In addition to rule application and vehicle handling, strategic skills are seen as vital for proficient driving. Examples of these are aligning life-level goals with travel purposes, reflecting on the effects of own driving behavior for traffic safety, planning and adjusting the route according to travel purposes. Throughout Europe (Hatakka et al. 2002; Peräaho et al. 2003) the learning outcomes for driver training have been changed accordingly, which has had major implications for educational designs and driver exams. The content of driver exams, both regarding the practical and theoretical exams, has been changed. In Dutch practical exams, test takers are expected to choose their routes independently and demonstrate self-reflection on the quality of driving (Vissers et al. 2008). In theory exams, attention is paid to the measurement of higher order skills, specifically hazard prediction in traffic, in addition to the application of traffic rules in meaningful traffic situations.

Second, technological assets for item construction, presentation and delivery have developed fast during the last decade (Almond et al. 2010; Almond et al. 2010; Drasgow et al. 2006; Drasgow and Mattern 2006). The use of computer-based online item authoring software enabled the use of innovative interactive items types, including hotspot, drag and drop, and timed response items, that more closely resembled the required responses as used in daily traffic. It is expected that both the change in content and in item types could have benefits for validity, which in the end improves driver safety.

In practice, theory exam delivery has changed throughout many European countries from linear classroom-based delivery into individual computer-based, where its content follows a European Directive on driving licenses (European parliament and council 2006). With regard to the Dutch theory exam, a first step in the transition of linear paper-based into non-linear computer-based exams involved the redesign of the

existing item collection, allowing online delivery in approximately 35 examination centers throughout the country.

The theory exam consists of two parts. The first part is a 25 item hazard perception test, which will not be discussed in this chapter. The second part, a 40 item subtest regarding traffic rules embedded in a traffic situation, is discussed here. The cut-score for passing this traffic rules test is a 35 out of 40 items correct score. The items typically consist of a specific traffic situation, presented in a picture, where the driver of a ‘leading car’, generally seen from the back, is about to carry out a traffic task (e.g. merging, turning, crossing, parking, overtaking) on a road section or an intersection. The candidate takes the perspective of the driver of this leading car.

Shortly after the introduction of the new computer-based delivery system a first evaluation study was carried out in 2014, in which the accessibility of a part of the redesigned items was reviewed, using the checklist described in Sect. 1.2 of this chapter. Results indicated possible sources of reduced accessibility (Zwitser et al. 2014). In general, most of the theory items were considered as accessible. Items that required a decision about the order of who goes first, second, third, fourth on an intersection using traffic rules and traffic signs, were seen as potentially less accessible. Other evaluative outcomes related to the stimuli used: in some cases the reviewers noted the use of too complex pictures, that did not correspond with the usual drivers’ perspective, e.g. a helicopter view of the car to determine dead angles around it.

2.4.2 Applied Modifications in the Response Mode of Theory Items

Following the results of the 2014 item screening, items that related to the application of Dutch traffic rules on intersections have been modified in order to improve the accessibility. Before 2014, test-takers for the Dutch theory exam were presented rule-of-way-at-intersections items in which they had to choose from verbal options (see left part of Fig. 2.3). The stem could be: “Which is the correct order to go?”. Test takers had to choose the correct option, which was a verbal enumeration of road users. In the 2014 study it was noted that the verbal response method could add extra extrinsic task load: keeping the verbal order in short term memory and compare it to the visual presentation of the road users in the item stimulus. Therefore the response mode for the rule-of-way-at-intersections items was changed into a direct visual response. As a response to the two respective item stems, test takers were either asked to drag corresponding rank numbers to the vehicle to indicate right of way order or to drag a check symbol towards the road user in the picture, who can go first at the intersection.

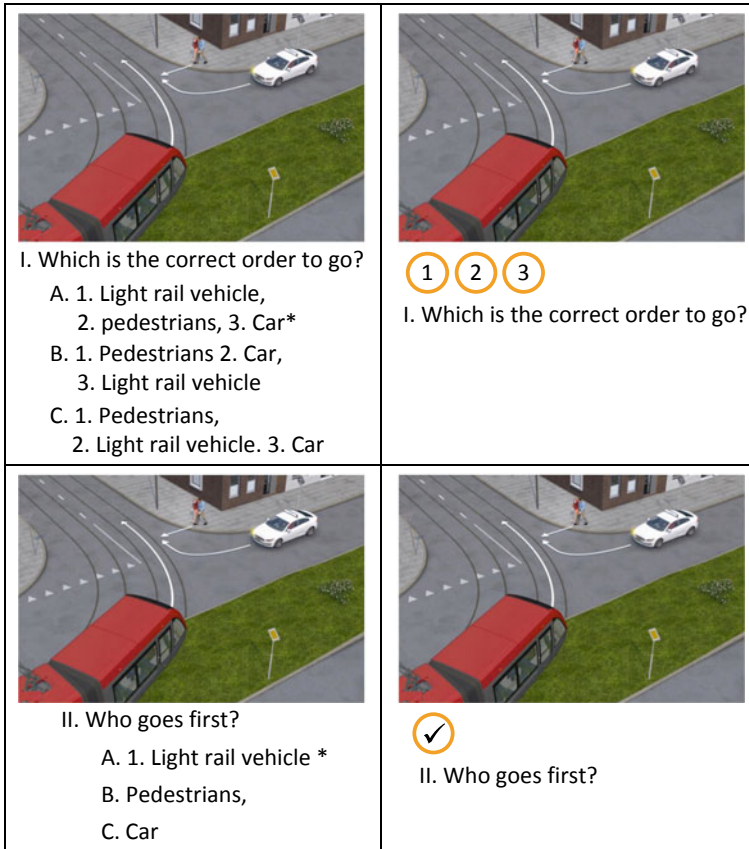


Fig. 2.3 Two items adapted form verbal into visual response mode

2.4.3 Psychometric Indications of Accessibility Improvement?

In studies regarding item modification, no clear cut indications are given of what counts as psychometric proof of improvement in item accessibility. Following prior studies of Beddow et al. (2013) and Beddow et al. (2011) we make the following argument.

From an accessibility perspective, a decrease in item difficulty after a modification is acceptable when it is caused by a reduction in extrinsic task load. In this case, the modified item had been more difficult for the wrong reasons, for instance because it required access skills unrelated to the task for the intended target group or because of inconsistencies in item presentation. Oppositely, a decrease in difficulty after modification would be less acceptable, when caused by a parallel reduction of intrinsic task load, e.g. for instance because certain knowledge is no longer needed to

respond correctly to the item. Vice versa, an increase in item difficulty is acceptable as long as there has been no increase in extrinsic task load, and as long as the increase in difficulty is balanced by other items in the intended test.

Ideally, item discrimination will at least remain the same or preferably increase after modification. The reduction of extrinsic task load can be expected to come with a more clear coverage of the target skill.

In the current study, the reasoning above was used, to investigate changes in item response format and their effect on item difficulty and item discrimination. The leading research question was:

To what extent are differences in response format, i.e. verbal or visual, related to item difficulty and item discrimination, when item type and content have been accounted for?

2.4.4 Item Selection

In this study, only items regarding the rule of way at intersections, involving three or four road users were taken into consideration, because the verbally presented orders within these items were expected to cause extrinsic task load.

For intersection items with three or four road users, including the leading car perspective, three different item stems have been used in either response mode. (1) “Which is the correct order to go?” (2) “Who goes first?” (3) “To whom do you have to give way?”

2.4.5 Data Collection

To enable the estimation of item parameters for the subset of chosen items we used the test data of 345 versions of the Dutch theory test, administered in the period between fall 2015 and spring 2018. The test data pertained to 294.000 test takers. Due to the fact that new items were added to and old items were removed from the item bank, the number of versions in which an item appeared, varied between 1 and 344, with a median value of 5 versions.

The test data did not contain data from accommodated versions of the theory test, i.e. individually administered versions, read aloud by test assistants.

2.4.6 Data Analyses

For the purpose of this study the test data of 107 intersection items were used, in which three or four road users arrive at the intersection.

Table 2.5 Prediction of Item difficulty (*Beta*) by item stem type and response mode in a multiple regression analysis

Variables in equation	<i>B</i>	<i>SE</i>		<i>t</i>	Sig.
Constant	−.22	.12		−1.89	.06
Stem: Which is the correct order to go?	.34	.17	.19	1.99	.05
<i>Excluded variables</i>					
Response mode: verbal (0)- visual (1)			.028	.197	.84
Stem: Who goes first?			−.043	−.35	.73

Multiple $R = .19$; Df regression = 1; Df residual = 105

Using Dexter, an R-application for Item-response calibrations (Maris et al. 2019), necessary item parameters have been estimated. A one parameter model was used to estimate beta-values. Mean item-rest correlations (Rir-values) were calculated for each item across the versions, where it had appeared. Using multiple regression analyses the predictive effects of two item features for item difficulty and item discrimination were assessed: item stem (three types, see Sect. 2.3.3) and response mode (verbal vs. visual).

2.4.7 Results

Multiple regression analyses with item difficulty (Beta-parameter) as a dependent variable (see Table 2.5) revealed that items in which test takers had to determine the correct order to go on the intersection were more difficult than items with a different stem, where the first to go was to be determined ($Beta = .19$, $p = .05$). Response mode was not associated with item difficulty ($Beta = .028$, $p = .84$)

Multiple regression analyses with item discrimination (item-rest correlation) as a dependent variable (see Table 2.6) showed that response mode was significantly associated with item discrimination ($Beta = .04$, $p = .00$). Items with a visual (drag and drop) response had higher item-rest correlations than items with a multiple choice verbal response. The type of stem used was not related to item discrimination.

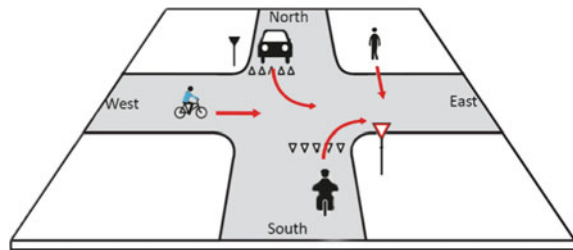
In Fig. 2.7 the item parameters for 16 paired items are displayed in a scatterplot. These pairs involve identical tasks and traffic situations, but differ in response type (visual response vs. verbal mode). It can be noticed from the upper part of Fig. 2.7 that for most of the items the change in difficulty is relatively low, after response mode change, with one clear exception. “Item A” in Fig. 2.4 is clearly more difficult in the visual response mode (beta-parameter = .13) than in the verbal mode (beta = −2.29). The content of this item is schematically depicted in Fig. 2.4. In order not to expose the original items, the original picture has been re-drawn. Item A involves four road users approaching a 4-way intersection. The traffic signs mean that participants on the East-West road have right of way vis a vis the participants in the North-South road. A pedestrian on the sidewalk North-left crossing into south direction, a car

Table 2.6 Prediction of Item discrimination by item stem type and response mode in a multiple regression analysis

	<i>B</i>	<i>SE</i>	<i>Beta-weight</i>	<i>t</i>	<i>Sig.</i>
Constant	.21	.01		32.40	.000
Response mode: verbal (0)- visual (1)	.04	.01	.35	3.76	.000
<i>Excluded variables</i>					
Stem: Which is the correct order to go?			-.17	-1.24	.218
Stem: Who goes first?			.06	.54	.588

Multiple *R* = .35; *Df* regression = 1; *Df* residual = 105

Fig. 2.4 Sketch of “item A”

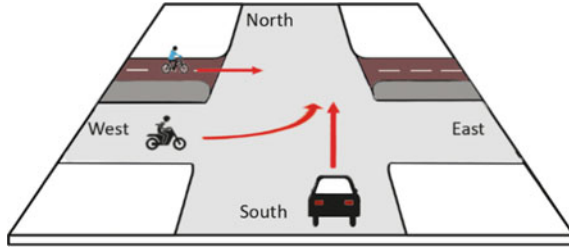


“Who goes first?
 A. bicycle
 B. motorbike
 C. car

turning left from the north lane, a motorbike from the south lane turning right and a cyclist from the west crossing in east direction. The question is who goes first out of four road users. The higher difficulty in the visual mode is most likely explained by the fact that in the verbal mode, the pedestrian is not mentioned in the response options. For pedestrians, the general rule of way does not apply. It only applies to drivers or riders of a vehicle. Because the pedestrian is not mentioned as an option, the test takers do not need to take into account this traffic rule. In the visual response mode however, the pedestrian needs to be considered, in order to indicate the road user who goes first, which increases the intrinsic task load.

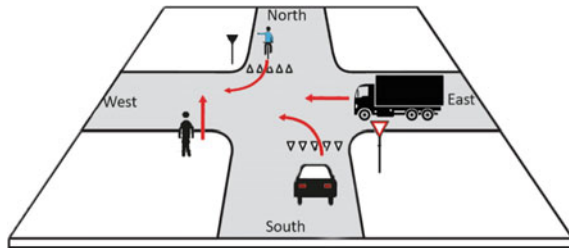
In the bottom part of Fig. 2.7, the Rir-values for 16 pairs of identical items, that were changed from a verbal into an visual response mode, are displayed in a scatterplot. In general it can be noted that most items have higher Rir-values after the adaptation of the response mode from verbal into visual. Item A, described above, and B and C have the highest increase in Rir-value. For item A the extra rule regarding the pedestrian, that needs to be considered in the visual response mode, will probably lead to an increase in possible options, contributing to the higher Rir.

Item B, the content of which is depicted in Fig. 2.5, shows a clear increase in difficulty (from 1.07 to 1.70), where at the same time the Rir-value increases (from .13 to .23). In item B three road users arrive at a non-signalized 4-way intersection,



Which is the correct order to go?
 A. 1 moped, 2 car, 3 motorbike
 B. 1. car, 2. moped, 3 motorbike
 C. 1. moped, 2. motor, 3 car

Fig. 2.5 Sketch of “item B”



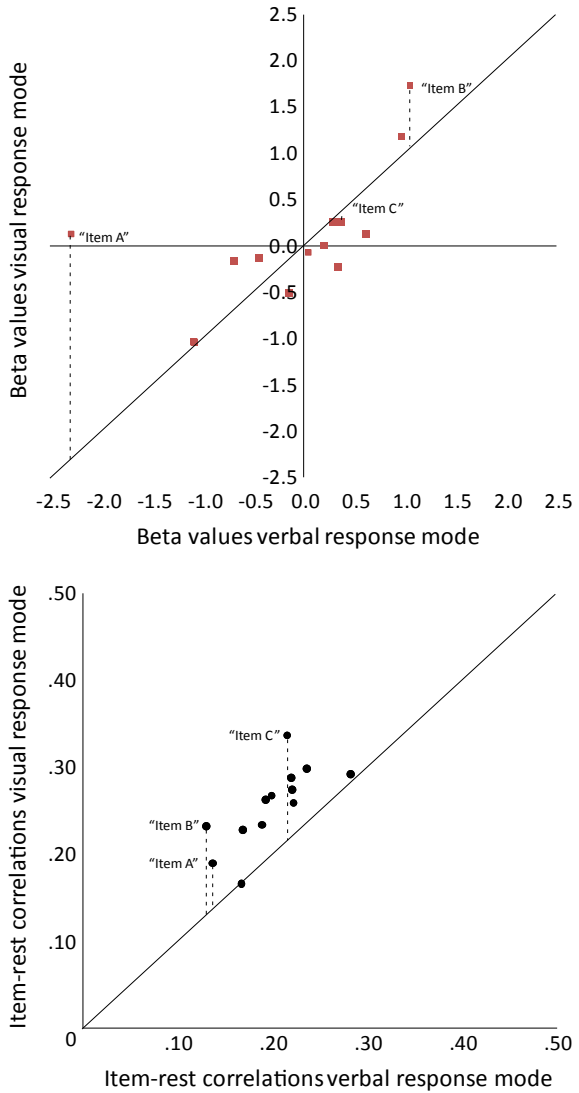
Who goes first?
 A. Bicycle
 B. Truck
 C. Pedestrian

Fig. 2.6 Sketch of “item C”

including a motorbike and a cyclist from the same direction (west), where the motorbike intends to make a left turn north, and the cyclist intends to cross eastwards. A car intends to cross from the south lane to the north lane. The explanation of the increased Rir is very likely the increased number of possible orders to go in the visual mode compared to the verbal mode. However, the Rir-value is still below what can be considered as appropriate.

Item C (depicted in Fig. 2.6) has the largest increase in Rir (from .22 to .34), but interestingly enough a relatively stable difficulty value (Beta decrease from .35 to .25). In this item, four road users arrive from four different directions at a 4-way intersection. The traffic signs mean that drivers and riders on the East-West road have right of way vis a vis drivers and riders in the North-South road. A car from south intends to turn left towards the west lane, a pedestrian is crossing from south to north on the left sidewalk, a cyclist turns right from the north lane into the west lane, and a truck crosses from the east lane towards the west lane.

Fig. 2.7 Item difficulty (betas) and item rest correlations for item pairs in a verbal and visual response mode



Both in the visual and verbal mode all applicable traffic rules need to be applied to arrive at the correct answer, although the car is not mentioned in the options. In this case, there is a possible exclusive response mode effect, in which the visual response takes away construct irrelevant task load, causing an increase in item discrimination (Fig. 2.6).

2.5 Discussion

In this chapter we presented a framework for design and review of accessible assessment tasks. We deliberately use the term “assessment task” to refer to tasks with any size, meant to be used in an assessment, that can have any form, be it a test or a performance assessment. Access in the context of educational testing refers to the opportunity for a student to demonstrate proficiency on a target skill. The framework has three basic elements: (1) cognitive load; (2) the test taker task solution process; and (3) parts of the assessment task presentation.

First, using the concept of cognitive load (Sweller 2010), we contended that each assessment task, intended to measure a target skill such as a mathematical word problem or solving an item about traffic rules on an intersection has a certain amount of intrinsic task load and extrinsic task load. Improving accessibility basically means optimizing the amount of intrinsic task load and minimizing the amount of extrinsic task load. It is up to the test designer to specify, what task elements are considered as intrinsic task load, and what comprises acceptable levels of intrinsic load for the target group test taker. In addition, elements that are seen as unavoidable extrinsic task load needs to be specified.

Second, using insights from cognitive laboratory studies we proposed to take the test takers task solution processes in mind in such a way as to facilitate these processes in assessment task design. In our framework, five general mental processes were considered that test takers go through when solving a problem. These entailed: (1) orientation on the task purpose and requirements, (2) comprehending task information, (3) devising a solution, (4) articulating a solution and (5) self-regulation including planning, monitoring and adjustment of the solution process. The fifth part of the process model was added by applying self-regulation theories on task execution (Winne 2001). Depending on the target domain (mathematics, science, reading, driving a car), these task solution processes may be broken down into further detail and specified for the skills content at hand. Additionally, test takers may make short cuts and arrive automatically at solutions, because sub processes are automated or take place unconsciously.

A third element of our model pertains to the parts of the assessment task presentation. We distinguished the assignment, task information, facilitation of responding, and item lay-out and test navigation. Again, these parts can be broken down further into subcategories or elaborated further. For instance, group tasks with collective outcomes and individual test taker contributions are not mentioned in our framework.

Combining these three elements we presented an approach for item design and screening, summarized by means of a checklist. Principles of existing instruments, such as the Test Accessibility and Modification Inventory (TAMI, Beddow et al. 2008) were applied in the checklist. In the checklist we concentrated on removing or avoiding extrinsic task load from assessments tasks, without changing the level of intrinsic task load. The idea behind it was that part of the inaccessibility is caused by the way assessments tasks are presented.

We reported a study in the context of theoretical driving exams, specifically regarding the application of rules of way on traffic intersections. This is a relatively complex skill, because Dutch traffic highway code has many different rules and signs that regulate the right of way on intersections. The transition of paper-based exams towards individually administered computer-based exams came with some item modifications after a first item review study (Zwitser et al. 2014). One of these modifications was the change of response mode regarding intersection items from verbal options into direct visual (drag and drop) response in order to mean to improve item accessibility. Using our framework we investigated whether this modification, meant to reduce the reading load and the necessity to change from visual orders to verbal orders, had effects on item difficulty and item discrimination. Reasoning from our accessibility framework, it would be desirable that a change in response mode would not affect the difficulty, but it should affect item discrimination. Reducing extrinsic task load should result in an improved target skill representation, which can be expected to show in an increase in item discrimination parameters.

The study including 107 items showed that the response mode mattered to item discrimination, as measured by the corrected item total correlation, although the effects were not spectacular. Regression analyses showed that items with a visual response mode discriminated better than those with a verbal response mode, when item stem differences had been accounted for. In addition we found that for items that address identical intersection tasks and that had identical stems, those with a visual response types discriminated significantly better than those with verbal response types. At the same time response mode was not significantly related to item difficulty, as measured by the IRT beta coefficient. There were a few exceptions for this result. We found that for one item the changed response type resulted in higher difficulty. This result could be explained by the increase of the number of traffic rules that needed to be applied to arrive at an correct answer. In the verbal option mode, one of the traffic rules did not need consideration, because the road user was not mentioned in the options.

This finding suggested that modifying items in order to improve accessibility needs to be done in combination with a look at the intrinsic task load at the same time. In follow-up research we intend to develop predictive models for item difficulty and item discrimination that combine intrinsic and extrinsic task features. In follow-up research it is our aim to identify task features that comprise intrinsic task load and task elements that comprise unavoidable and irrelevant task load. By specifying these task features in advance, it can be determined which elements of the assessment tasks need some kind of access skill and which address target skills.

To conclude this chapter some theoretical, practical and methodological limitations and challenges can be mentioned. First, it is agreed upon that accessibility is not seen as a static test property but instead, as the result of an interaction among test features and person characteristics that either permit or inhibit student responses to the targeted measurement content (Kettler et al. 2009). In the current study on traffic intersection items, the data did not allow us to look for specific interactions between test takers features and item features. A research question could have been whether

test takers with reading or attention problems benefitted more from the changed response modes than other test takers.

Second, a validity issue is to which extent changes in accessibility affect extrapolation of test results to target situations (Kane 1992, 2004). Do changes in accessibility correspond with accessibility in the target situations? The question is, which are absolutely unavoidable sources of extrinsic task load? What if the target situations contain comparable sources of extrinsic task load? Applying to our example regarding driving on intersections, there may be all kinds of distracting information, e.g. adverse weather conditions, unexpected maneuvers of other road users, crying kids in the back of the car. To extrapolate from the test towards performance in the practice domain, we probably need a sound definition of the target skill and the target situation. In this example, drivers need to be situationally aware, prepared to notice changes in the traffic situation. A theory test as described, with a picture and a question like “who goes first”, cannot account for these situational changes. It was not designed for the purpose of taking into account situational awareness (Crundall 2016; McKenna and Crick 1994). It would however be possible to design an interactive simulator-driven test with scenarios that presents tasks, where application of traffic rules is part of a larger set of target skills, including determining whether it is safe to cross, taking into account both the rule and the speed and position of other road users.

This brings us to a final, much broader issue, involving the question how to map the development of access skills and target skills over time, using developmental trajectories and what this means for designing construct relevant and accessible assessments, that inform learning processes (Bennett 2010; Pellegrino 2014; Zieky 2014). How do target skills develop over time? To what extent do skills develop into access skills for other skills, such as reading skill for math? How do these skills get intertwined to form new target skills? Which intermediate learning progressions take place? In order to inform learning design for students with different needs in different developmental stages, we need to find answers to several design questions: what constitutes necessary intrinsic task load to cover target skills? What goes on in test takers during task solution? What levels of access skills can be assumed and which sources of extrinsic task load are unavoidable or unnecessary? In order to find answers to these questions in the future, a combination of approaches provide a more comprehensive framework, including an evidence centered design of assessments (Mislevy and Haertel 2006) using cognitively based student models (Leighton and Gierl 2011) elaborated in specified item models (Gierl and Lai 2012) to be tested in multiple educational contexts.

Acknowledgements I am grateful to Frans Kamphuis, Timo Bechger, and Angela Verschoor for the item calibration activities they have carried out.

References

- Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*, 219–234.
- Almond, P., Winter, P., Cameto, R., Russell, M., Sato, E., Clarke-Midura, J., et al. (2010a). Technology-enabled and universally designed assessment: Considering access in measuring the achievement of students with disabilities—A foundation for research. *The Journal of Technology, Learning, and Assessment, 10*(5), 1–41.
- Almond, R. G., Steinberg, L. S., & Mislavy, R. J. (2010b). Enhancing the design and delivery of assessment systems: A four process architecture. *The Journal of Technology, Learning, and Assessment, 1*(5), 1–63.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.
- Beddow, P. A., Kettler, R. J., & Elliott, S. N. (2008). *Test accessibility and modification inventory*. Nashville, TN: Vanderbilt University.
- Beddow, P. A., Kurz, A., & Frey, J. R. (2011). Accessibility theory: Guiding the science and practice of test item design with the test-taker in mind. In S. N. Elliott, R. J. Kettler, P. A. Beddow, & A. Kurz (Eds.), *Handbook of accessible achievement tests for all students bridging the gaps between research, practice, and policy* (pp. 163–182). New York: Springer.
- Beddow, P. A., Elliott, S. N., & Kettler, R. J. (2013). Test accessibility: Item reviews and lessons learned from four state assessments. *Education Research International, 2013*, 1–12.
- Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research & Perspective, 8*(2–3), 70–91.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281. <https://doi.org/10.3102/00346543065003245>.
- Crundall, D. (2016). Hazard prediction discriminates between novice and experienced drivers. *Accident Analysis and Prevention, 86*, 47–58. <https://doi.org/10.1016/j.aap.2015.10.006>.
- Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology, 20*, 405–438.
- Deane, P. (2011). *Writing assessment and cognition*. Research Report ETS RR–11–14. Princeton: Educational Testing Service.
- Dolan, R. P., & Hall, T. E. (2001). Universal design for learning: Implications for large-scale assessment. *IDA Perspectives, 27*(4), 22–25.
- Dolan, R. P., & Hall, T. E. (2007). Developing accessible tests with universal design and digital technologies: Ensuring we standardize the right things. In L. L. Cook & C. C. Cahalan (Eds.), *Large-scale assessment and accommodations: What works* (pp. 95–111). Arlington, VA: Council for Exception Children.
- Drasgow, F., & Mattern, K. (2006). New tests and new items: Opportunities and issues. In D. Bartram & R. Hambleton (Eds.), *Computer-based testing and the internet* (pp. 59–76). Hoboken, NJ: Wiley.
- Drasgow, F., Luecht, R. M., & Bennett, R. (2006). Technology and testing. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 471–515). Washington, DC: American Council on Education/Praeger Publishers.
- European parliament and the council of 20 december 2006 on driving licences. (2006). Directive 2006/126/EC (recast). Official Journal of the European Union (2016), December 30, L 403/18- L 403/13. Retrieved at: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006L0126&from=en>.
- Gierl, M. J., & Lai, H. (2012). The role of item models in automatic item generation. *International Journal of Testing, 12*(3), 273–298.

- Gianns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*(6), 511–525.
- Hatakka, M., Keskinen, E., Gregersen, N. P., Glad, A., & Hernetkoski, K. (2002). From control of the vehicle to personal self-control: Broadening the perspectives to driver education. *Transportation Research Part F, 5*, 201–215.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*(3), 527–535.
- Kane, M. T. (2004). Certification testing as an illustration of argument-based validation. *Measurement: Interdisciplinary Research and Perspectives, 2*, 135–170.
- Ketterlin-Geller, L. R. (2005). Knowing what all students know: Procedures for developing universal design for assessment. *Journal of Technology, Learning, and Assessment, 4*(2), 1–23.
- Ketterlin-Geller, L. R. (2008). Testing students with special needs: A model for understanding the interaction between assessment and student characteristics in a universally designed environment. *Educational Measurement: Issues and Practice, 27*(3), 3–16. <https://doi.org/10.1111/j.1745-3992.2008.00124.x>.
- Kettler, R. J., Braden, J. P., & Beddow, P. A. (2011). Test-taking skills and their impact on accessibility for all students. In S. N. Elliott, R. J. Kettler, P. A. Beddow, & A. Kurz (Eds.), *Handbook of accessible achievement tests for all students bridging the gaps between research, practice, and policy* (pp. 163–182). New York: Springer.
- Kettler, R. J., Elliott, S. N., & Beddow, P. A. (2009). Modifying achievement test items: A theory-guided and data-based approach for better measurement of what students with disabilities know. *Peabody Journal of Education, 84*, 529–551.
- Leighton, J. (2017). *Using think-aloud interviews and cognitive labs in educational research*. New York, NY: Oxford University Press.
- Leighton, J. P., & Gierl, M. J. (2011). *The learning sciences in educational assessment: The role of cognitive models*. Cambridge, UK: Cambridge University Press.
- Mace, R. (1997). *The principles of universal design* (2nd ed.). Raleigh, NC: Center for Universal Design, College of Design. Retrieved May 20, 2010, from http://www.design.ncsu.edu/cud/pubs_p/docs/poster.pdf.
- Maris, G., Bechger, T., Koops, J., & Partchev, I. (2019). DEXTER. Data management and analysis of tests, version 0.8.4. Manual Retrieved at: <https://cran.r-project.org/web/packages/dexter/dexter.pdf>.
- McKenna, F. P., & Crick, J. L. (1994). *Hazard perception in drivers: A methodology for testing and training*. TRL Contractor Report (313).
- Mislevy, R. J., & Haertel, G. (2006). Implications for evidence-centered design for educational assessment. *Educational Measurement: Issues and Practice, 25*, 6–20.
- Pellegrino, J. W. (2014). Assessment as a positive influence on 21st century teaching and learning: A systems approach to progress. *Psicología Educativa, 20*(2), 65–77.
- Peräaho, M., Keskinen, E., & Hatakka, M. (2003). *Driver competence in a hierarchical perspective: Implications for driver education*. Turku: Turku University, Traffic Research.
- Roelofs, E. C. (2016). Naar richtlijnen voor het ontwerp van toegankelijke toetsopgaven [Towards the design of accessible test items]. *Examens, 13*(3), 4–11.
- Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research, 75*(4), 457–490.
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*, 123–138.
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved [2-1-11], from the World Wide Web: <http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html>.
- Vissers, J., Mesken, J., Roelofs, E. C., & Claesen. (2008). New elements in the Dutch practical driving test: a pilot study. In L. Dorn (Ed.), *Driver behavior and training* (Vol. III, pp. 37–50). Hampshire: Ashgate Publishing Limited.

- Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed., pp. 153–189). Mahwah, NJ: Lawrence Erlbaum Associates.
- Winter, P. C., Kopriva, R. J., Chen, Ch S., & Emick, J. E. (2006). Exploring individual and item factors that affect assessment validity for diverse learners: Results from a large-scale cognitive lab. *Learning and Individual Differences, 16*, 267–276.
- Zieky, M. J. (2014). An introduction to the use of evidence-centered design in test development. *Psicología Educativa, 20*, 79–87.
- Zucker, S., Sassman, C., & Case, B. J. (2004). *Cognitive Labs*. Pearson Technical report. New York, NY. Retrieved at http://images.pearsonassessments.com/images/tmrs/tmrs_rg/cognitivelabs.pdf.
- Zwitser, R., Roelofs, E., & Béguin, A. (2014). *Rapportage van het onderzoek naar de psychometrische kwaliteit van de huidige CBR theorie-examens [Research into the psychometric quality of the current Dutch driving theory-exams]*. Arnhem: Cito.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

