1-2018

New approaches to selection system design in healthcare: The practical and theoretical relevance of a modular approach

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg

Jan CORSTJENS

New Approaches to Selection System Design in Healthcare:

The Practical and Theoretical Relevance of a Modular Approach

Filip Lievens

Singapore Management University, Singapore

Jan Corstjens

Ghent University, Belgium

Lievens, F., & Corstjens, J. (in press). New approaches to selection system design in healthcare: The practical and theoretical relevance of a modular approach. In: F. Patterson & L. Zibarras (Eds.). *Selection and Recruitment in the Healthcare Professions: Research, Theory and Practice*. Palgrave McMillan.

Abstract

This chapter presents a modular approach to healthcare selection system design. Contrary to the traditional holistic view on selection procedures, a modular approach highlights the components underlying selection procedures. Our framework identifies seven key design components of selection procedures (The stimulus format, contextualization, stimulus presentation consistency, the response format, response evaluation consistency, information source, and instructions) and reviews studies in the healthcare selection literature that compared the effect of these components on key selection outcomes. A modular approach allows (1) gaining insights into how the different components underlying selection procedures affect selection outcomes and (2) drawing conceptual similarities between components of different selection procedures. At a practical level, a modular approach permits developing a myriad of new selection procedures by "mixing and matching" different building blocks. We present two case studies and future research avenues to further illustrate these merits of a modular approach.

Keywords: Modular, Selection, Format, Scoring.

Prof. Dr. Filip Lievens

Lee Kong Chian Professor of Human Resources

Singapore Management University

Lee Kong Chian School of Business

50 Stamford Road

Singapore 178899

Singapore

Filip Lievens is the Lee Kong Chian Professor of Human Resources at the Lee Kong Chian School of Business of Singapore Management University. He is also an Honorary Professor at the University of Cape Town and a visiting professor at Ghent University, Belgium, where he also obtained his PhD in 1999. He has published over 160 articles in the areas of selection, assessment, and recruitment. His research interests deal with talent assessment, situational judgement, intra-individual variability, personality, and organizational attractiveness. His articles have appeared among others in the *Annual Review of Psychology, Journal of Applied Psychology, Personnel Psychology,* and *Journal of Management.*

Drs. Jan Corstjens

PhD Candidate

Ghent University

Faculty of Psychology and Educational Sciences

Department of Personnel Management, Work and Organizational Psychology

Henri Dunantlaan 2

9000 Ghent

Belgium

Jan Corstjens is PhD candidate at the Department of Personnel Management and Work and Organizational Psychology at Ghent University, Belgium. He obtained his Bachelor´s degree in Cognitive Psychology (2011) and his Master's degree in Work and Organizational Psychology (2012) from Maastricht University, the Netherlands. His research interests are positioned in the field of selection and assessment. Specifically, he focuses on selection procedures that assess academic and non-academic constructs, and how cognitive-motivational processes affect test performance.

1. Introduction

Traditionally, healthcare selection practices and research can be characterized by a "holistic" view on selection procedures. For example, prior reviews on the various selection procedures that can be used as part of healthcare selection (Patterson et al., 2016; Roberts et al., 2017) focus on ability tests, aptitude tests, references, interviews, personality inventories, situational judgement tests, etc. Clearly, such a holistic focus is relevant because this is how the selection procedures are typically used in practice.

Recently, Lievens and Sackett (2017) introduced a new approach to conceptualizing and examining selection systems, methods and processes. Specifically, they argued in favour of a modular approach to personnel selection. In a modular approach, a selection procedure is broken down into smaller components (aka "building blocks"). This means that a selection procedure is seen as an assemblage of such loosely-coupled and relatively independent building blocks. Examples of such building blocks are stimulus format (e.g., video-based vs. written way of presenting items) or instruction format (e.g., vague or more direct directions to take the test). These components can then be flexibly recombined to construct a wide variety of new "hybrid" selection procedures. Whereas Lievens and Sackett (2017) applied such a modular approach to the broader personnel selection field, the aim of this chapter is to present and apply a modular approach to the healthcare selection field.

The structure of this chapter is as follows. First, we introduce the essential characteristics of a modular approach and delineate its conceptual and practical merits. Next, we review research in healthcare selection that has already adopted a modular approach. At the same time, we propose avenues for future research to promote modular thinking in healthcare selection practices. In a third part, we present case studies that applied such a modular approach in a specific practical setting.

By the end of this chapter, you should have some understanding of and be able to explore further:

- The essentials of a modular approach to selection system design;
- The seven building blocks of a modular approach to selection;
- The conceptual and practical merits of a modular approach;
- How to apply a modular approach in healthcare selection research;
- How to use a modular approach to healthcare selection in practice.

2. What is a Modular Approach to Selection?

The term modular originated in the product design literature. In that literature, two schools are often distinguished: integrated architectures in which systems are seen as holistic "all-in-one" packages versus a modular perspective that breaks down systems into key loosely coupled components (aka "building blocks"). So far, the holistic "all-in-one" perspective has dominated the selection field.

Lievens and Sackett (2017) conducted an in-depth literature review and identified seven such key building blocks (aka the “Super Seven”). First, the stimulus format refers to the modality by which the test stimuli (e.g., information, questions, prompts) are presented to test-takers. For example, stimuli might be presented in a written (e.g., written verbal reasoning items), pictorial (e.g., facial pictures in an emotional intelligence task), or auditory (e.g., telephone interview questions) format; other options are dynamic audio-visual stimuli (e.g., video scenes or 3D animated scenes), videoconference (aka remote, online) interactions, and live face-to-face interactions.

Second, contextualization is defined as the extent to which test stimuli are embedded in a detailed and realistic context. In selection, some tests are decontextualized because no context is provided to test-takers. Examples are the classic cognitive ability tests. In low levels of contextualization**,** a situational keyword (aka tag) is added. An example is adding an

"at work" tag to existing personality items. A medium level of contextualization means that the context is described in terms of "who", "when", "where", and "why" (Johns, 2006). Examples are situational interviews and situational judgement tests (SJTs). Conversely, in high levels of contextualization, more detailed information is given about each of these aspects (e.g., a one-page description of the main characters). Examples include serious games or assessment centre exercises.

Third, stimulus presentation consistency refers to the level of standardization that interviewers/assessors/test administrators adopt in presenting test stimuli to test-takers. In other words, this factor refers to the degree to which procedural variations in presenting test stimuli across test takers are reduced. Lievens and Sackett (2017) made a distinction between three subcategories. Little standardization regarding the stimuli to be presented ("free stimuli"). Examples are interviews or role plays without question standardization. In the case of adaptive stimuli, there exist predetermined and standardized guidelines about the key stimuli to be presented but the administration of sub-stimuli depends on test-takers' responses to the previous stimuli. The highest level of stimulus presentation consistency is exemplified by fixed stimuli: All test-takers are presented the same or comparable stimuli in the same order (no matter how they respond to the stimuli) and progress in the same way.

Fourth, the response format denotes the modality by which test-takers are required to respond to test stimuli. Closed-ended response formats (multiple-choice or forced-choice response formats) in which possible response options are predetermined and prompted are frequently used. Candidates choose, rank, or rate the predetermined response options. The closed-ended response options might be text-based, pictorial, auditory, or video-based. Apart from closed-ended formats, open-ended (aka constructed) response formats have also been used. The same categories (textual, pictorial, audio, audio-visual, videoconference, and face-to-face) apply here as the ones discussed for stimulus format. For example, in recent times

candidates have submitted video resumes (Waung, Hymes, & Beatty, 2014), have answered test items via the webcam (Lievens, De Corte, & Westerveld, 2015), or have engaged in videoconference (remote) or live interactions with interviewers.

Fifth, response evaluation consistency refers to the level of standardization that evaluators adopt to evaluate test-takers' responses. Again, different subcategories can be distinguished. In unconstrained judgment, candidates are evaluated without having pre-established answers or a scoring key. Conversely, calibrated judgment signifies that evaluators are trained to use pre-established answers and/or evaluative standards when evaluating candidates, as is often the case in scoring interview answers, essays, and role-plays. Automated scoring represents the category highest in terms of response evaluation standardization.

Sixth, information source refers to the individuals responding to the test stimuli. The typical distinction here is between the focal person and reports by other informants. These persons should be well acquainted with the focal person and motivated to share job-related information about him/her. Examples of such others are co-workers, supervisors, or direct subordinates. The effects of using different judges to evaluate candidate performance also fall under this component.

Finally, instructions denote the extent to which directions are made explicit to test-takers about which perspective to take to respond to test stimuli. A distinction can be made between general (weaker e.g., "rate yourself on the following statements") and specific (stronger) instructions (e.g., with clearer directions on how to solve the item/task). Table 1 summarizes these seven building blocks, their definitions, and subcategories (see also Lievens & Sackett, 2017).

INSERT TABLE ONE*

3. What are the Benefits of a Modular Approach to Selection?

A modular approach has a lot of merit for healthcare selection. Generally, Lievens and Sackett (2017) argued that a modular approach has various conceptual and practical benefits. Two conceptual benefits of a modular approach are of most importance. First, a modular approach leads to greater insight into the workings of each of the separate components because the isolated impact of these components is examined on key selection outcomes. For example, when one focuses on a holistic selection procedure such as a selection centre, it is difficult to determine why it leads to valid predictions of future performance. It might be because the assessors are well trained, the exercises are contextualized, or because the response format is not closed-ended. Conversely, if one examines the effect of one component such as response format (while keeping all the others factors constant), one might determine whether an open-ended (as compared to a closed-ended) response format leads to better predictions.

Second, a modular approach creates more integration and cross-fertilization across different selection procedures because these components cut across various selection procedures. Returning to the selection centre example above, suppose one finds that an open-ended response format leads to better predictions and thus higher validity, such insights might inform not only selection centre design but a variety of other selection procedures such as work samples, simulations, etc. Other examples are that there exist different literatures per selection procedure. Think, for example, about the panel interview literature and the selection centre literature. By focusing on building blocks (e.g., response scoring consistency) that underlie both selection procedures, one generates insights for both literatures and accordingly fosters integration and theoretical connectivity among them.

As a key practical benefit, a modular perspective permits developing a myriad of new selection procedures by "mixing and matching" different building blocks. That is, one might design a new selection procedure by changing one or more building blocks of an existing one. For example, one might invest in higher levels of response scoring consistency or more contextualization when designing an interview. Similarly, one might increase the level of contextualization of a personality inventory. Such changes might be made to improve reliability, validity, applicant perceptions, or reduce costs and gender, ethnic or socioeconomic differences in test scores, thereby widening access to healthcare education for minority groups. This ability to change building blocks leads to increased agility in (re)designing selection procedures, which serves as a catalyst for innovation and change.

4. Review of Healthcare Selection Research Using a Modular Approach

Two types of studies can typically provide data on the effects of how each of these building blocks affect key selection criteria such as reliability, validity, minority group test score differences, or applicant reactions. First, there are primary studies that conduct a comparative evaluation of predictor method factor choices. In a prototypical primary study, one building block (e.g., stimulus format, Chan & Schmitt, 1997) might be manipulated, with other aspects (i.e., test content, other method factors) being held constant. Second, moderator studies in meta-analyses can shed light onto the effects of building blocks. Such meta-analytic evidence has the advantage of being more cumulative. Yet, this also comes with a price because other potentially important factors are often not controlled for.

As compared to the personnel selection field in general (e.g., Christian, Edwards, & Bradley, 2010), we are not aware of such meta-analyses in the healthcare selection domain. However, so far, various primary studies in healthcare selection have already adopted a (partially) modular approach. Below, for each building block we present some examples of

these studies when available, without having the intention to be comprehensive in our coverage. In addition, we outline various avenues for future research per building block.

4.1 Effect of the Stimulus Format

The choice of modality by which test stimuli (information, questions, prompts) are presented to test-takers might affect important selection outcomes. For example, varying stimulus format might change the selection method's construct saturation (i.e., the strength of the relation with constructs such as cognitive ability or personality), criterion-related validity, minority group differences in test scores, and applicant reactions. Evidence from the broader selection field attested to this for selection interviews (videoconference vs. face-to-face interviews, Van Iddekinge, Raymark, & Roth, 2005) and situational judgement tests (SJT, Lievens & Sackett, 2006).

In the healthcare selection domain, an example of a primary study varying the stimulus format is one conducted by Lievens and Sackett (2006) in the context of the healthcare admissions exam in Flanders. This exam includes an SJT to measure students' knowledge of interpersonal and communication skills. In one cohort of candidates, the SJT was presented in a video-based format, while in a subsequent albeit similar cohort the stimulus format was changed to a textual one. All other components of the SJT were held constant. This allowed the researchers to study how these different stimulus formats affected important outcomes while the rest of the procedure was held exactly the same. Results revealed that the SJT with the video-based stimulus format evidenced a significantly lower correlation with cognitive measures, as well as significantly higher validity for predicting performance on interpersonally oriented courses later on. Moreover, the SJT with the video-based stimulus format showed significant incremental validity over other selection tests in the healthcare admission exam (i.e., a cognitive ability and a work sample test), whereas the

textual version did not. Lastly, applicant perceptions were also more positive for the video-based format compared to the written SJT.

Dore, Reiter, Kreuger, and Norman (2017a, 2017b) adopted a similar approach as the Lievens and Sackett study, although they administered two different stimulus formats in one and the same cohort of medical school applicants. The SJT contained items with either a written or a video-based stimulus format. Predictive validity showed a similar pattern as in the Lievens and Sackett study: the video-based format was significantly predictive for interpersonally oriented criteria gathered during a national licensure exam, whereas the written component was not. Other studies where stimulus format is varied could potentially adopt the same approach. For example, Husbands et al. (2015) examined the reliability and construct-related validity of an integrity-based SJT by combining scores of three different stimulus formats (written format, video-based format, and a written verbatim transcript of the video-based format). While the combined SJT evidenced satisfactory reliability and construct-related validity with integrity measures, applying the modular approach could have addressed if these metrics would differ for each stimulus format.

In the future, technological advancements that enable including novel stimulus formats could also be used and studied in a healthcare selection context. For instance, 3D or avatar-based multimedia environments have advanced to a point that they can represent high fidelity social interactions. Use of such high-fidelity technology might further increase favourable applicant reactions and criterion-related validity, and reduce minority group differences in test scores (Weekley, Hawkes, Guenole, & Ployhart, 2015). Translating this research to healthcare selection has the advantage of identifying a stimulus format or a range of stimulus formats that a healthcare school for example finds most beneficial in fulfilling its objectives (e.g., in terms of validity improvement, ensuring favourable applicant perceptions, widening access to minority populations).

4.2 Effect of Contextualization

In healthcare selection, methods such as selection centres or situational judgement tests often frame stimuli such as questions, information, or cues in a context that closely resembles practice (e.g., a physician's practice). Simulations with standardized patients are a prime example of such a contextualized selection procedure. Standardized patients are commonly trained lay persons that have been instructed to play the role of a patient and present various disease symptoms to the student physician during a simulation (Barrows, 1993; Cleland, Abe, & Rethans, 2009). In the healthcare selection literature and the selection literature more broadly though, studies that zoom in on contextualization as a distinct component of a selection method are sparse.

Future research could focus on identifying and manipulating different levels of contextualization which allows to study the effects of these differences on test scores and selection outcomes. Along these lines, Johns (2006) differentiated between omnibus (who, where, when, why) and discrete descriptions(specific details about social, task, and physical aspects of the context) as important dimensions of context. Besides varying the amount of contextualization provided, test stimuli could also be embedded in different types of contexts. For example, an entry-level student selection system could embed items in either a familiar context such as a classroom team project setting or opt to frame items in unfamiliar situations such as a future physician's daily patient encounters (see also case study 2 below). Clearly, investigating contextualization as a selection method component might have important implications for practice. For example, if scores on test items that strip away context show similar validity coefficients as their more highly contextualized counterparts, investments in such seem to pay off more.

4.3 Effect of Stimulus Presentation Consistency

Selection test developers often have to deal with making trade-offs between presenting stimuli to test takers in a highly fixed (and arguably more artificial) manner versus

presentation of stimuli in a more open-ended or adaptive format. The former increases test standardization, whereas the latter ensures a higher amount of real-world likeness and therefore more realistic presentations of a practicing physician's daily encounters. To maximize the selection method's potential to predict a physician's career success, developers attempt to create simulations that mimic these actual situations encountered in medical practice as closely as possible. Returning to the simulated patient example outlined before, the psychometric disappointments that arose when presenting stimuli only in a fixed format in patient management problems, and the fact that real-life patient-physician interactions are a free inquiry process led to the development of patient-physician simulations with simulated patients (Vu & Barrows, 1994).

Although we were not able to identify primary examples in the healthcare selection literature that directly compared different stimulus presentation consistency formats within the same selection method, we believe such comparisons might shed further light on important selection outcomes (criterion-related and construct validity, gender/ethnicity differences) and might inform schools where in the healthcare education pathway each format would be most suited to maximize the potential of selecting the best candidates (e.g., entry-level selection versus selection for specialty training). Future studies examining the effects of stimulus presentation consistency can also vary the signals or cues that are presented by simulated patients to be able to differentiate between students' swiftness in their ability to pick up on important disease signals. For example, a simulated patient can first present symptoms of the flu in a subtle nonverbal way (e.g., sniff), and when this is not picked up by the student the patient can move on to expressing a more obvious prompt (e.g., cough) and finally a verbal clear prompt (e.g., complain about a sore throat and runny nose)[1].

[1] Note that these examples can also be regarded as a manipulation of the instruction strength (more specific instructions versus more vague global ones, see below).

4.4 Effect of the Response Format

"Life is not about multiple choice" is an often-heard critique on selection methods with closed-ended response formats. Returning to the example outlined above, written patient management problems provide students with a written patient description and a list of options to choose from related to eliciting medical information about the patient. Such cued and closed-ended simulations are cost-effective to administer to large groups. Yet, they do not simulate the critical patient-doctor free-inquiry process (Vu & Barrows, 1994). Patient-physician simulations with standardized patients have an open-ended constructed response format and therefore can be thought to more closely resemble a physician's actual practice.

In one of the studies in healthcare selection that directly compared different response formats, Norman and Feightner (1981) comparatively evaluated medical students' performance on closed-ended patient management problems versus open-ended patient-physician simulations with stimulated patients. Clerkship students had to partake in periodical evaluations that presented them with either written patient management problems (close-ended format) or simulated patient encounters (open-ended format). The patient management problems were structured around a case (a patient with an illness) that was followed by a list of options the student could select to cue the patient for more information which then ideally should result in the ability to correctly diagnose the patient and solve the case. When comparing performance of the clerkship students on these written patient management problems with their performance on simulated patient cases, the students selected much more options to elicit information from the patient in the close-ended format than they would ask the simulated patient questions in the open-ended free-inquiry process. The authors concluded that the test format with fixed stimuli introduced error variance resulting from the test format and that the higher levels of standardization thus came at a cost in comparison to the less standardized free inquiry process with simulated patient formats. It should be noted, though, that in this study (apart from the response format) stimulus format was also

manipulated (written case vs. actual interactions), which makes it difficult to draw strong conclusions about the response format effects.

Research in the broader personnel selection literature further evaluated whether open-ended response formats (that are arguably more life-like than closed-ended formats) also translated into higher predictive validity. For interpersonal skill measurement, it seems the constructed response format improves criterion-related validity over closed-ended formats for predicting interpersonal criteria (Lievens, De Corte, & Westerveld, 2015, see also case study example 1 below). Importantly, other benefits of the open-ended constructed response format seem to be a reduction of cognitive load, lower ethnicity differences in test scores, and more favourable applicant perceptions (for a review, see Lievens & Sackett, 2017).

Future research in healthcare selection might attempt to echo these investigations from the broader personnel selection literature and seek to comparatively evaluate validity of different forms of response formats while keeping the remaining test content constant. For example, computerized assessments of clinical and interpersonal skills can record respondents' constructed answers with a webcam and compare them with answers provided in multiple-choice formats.

4.5 Effect of Response Evaluation Consistency

The validity of many selection methods is likely influenced by the level of standardization adopted in the scoring method. Examples in the domain of objective structured clinical examinations (OSCEs, Harden, Stevenson, Downie, & Wilson, 1975) illustrate how this component has already been investigated in healthcare selection utilizing a modular approach. Examinations of response evaluation consistency in OSCE stations showed that interrater agreement was higher when checklists were used than when rating scales were used. It was argued that this could be due to checklists being more concrete and objective definitions of potential behaviours to be demonstrated by candidates (Van der

Vleuten & Swanson, 1990). Conversely, interstation reliability was higher for more global and arguably more subjective rating scales, suggesting that across situations global judgments seem to provide more reliable estimates of candidates' abilities than specific ones (Swanson & van der Vleuten, 2013).

Another example of a recent study that can serve as an example here deals with an investigation of the effects of situational judgement test scoring method on internal reliability estimates and ethnic differences in test scores (De Leng et al., 2017). In this study conducted in the context of medical selection in the Netherlands, an extensive set of scoring methods were compared. The authors compared the psychometric properties of scoring methods that used raw consensus scoring (scores are awarded when test takers' ratings resemble the ratings provided by a reference group), versus standardizing techniques (applying the same comparison technique as raw consensus scoring after applying a within-person standardization of the scores to a scale with mean 0 and *SD* 1), and dichotomizing scores (awarding scores when test takers' ratings are on the same end of the scale as those of the reference group). By varying the response evaluation consistency component, the authors showed that ethnic differences in test scores could be influenced by the choice of scoring method. Raw consensus scoring appeared the most suited technique to maximize test score reliability and simultaneously minimize differences in test scores resulting from ethnic descent.

A modular approach might be especially beneficial for future research in further improving interview reliability and validity in healthcare selection. Different scoring rubrics can be experimented with. In healthcare selection (and the broader domain of personnel selection), the admission interview has been plagued by mixed findings regarding its predictive validity (Patterson et al., 2016). This seems to be in part related to the level of standardization adopted in presenting stimuli (high vs. low levels of question

standardization), but also to the level of standardization taken when evaluating responses (scoring). For example, scoring the performance of healthcare students can be conducted in an absolute (criterion-referenced) or a relative fashion (norm-referenced). Uijtdehaage, Doyle, and Parker (2011) compared these two scoring schemes in multiple mini interview (MMI) stations and found that norm-referenced scoring removed the ceiling effect that was apparent in the criterion-referenced scoring system, thereby enabling assessors to differentiate between the most highly competent students.

4.6 Effect of Information Source

Apart from the distinction between methods that require different information sources to provide information about the candidate (e.g., self-reports vs other-reports), one can also study the effects of various rating sources who evaluate a candidate's performance within one selection method. For example, in order to reduce costs and time investments from experienced practicing physicians, several studies in the context of OSCEs comparatively evaluated different rating sources to examine whether rating source affects the reliability of the scores. The typical comparison has been between experienced physicians and non-faculty experts or standardized patients. Although experienced physicians would appear to be the most qualified group to rate and evaluate the performance of candidates, when given sufficient training the ratings provided by non-faculty experts or standardized were found to be equally reliable and accurate (Swanson & van der Vleuten, 2013; Vu & Barrows, 1994). In addition, the number of assessors seems to matter: optimal reliability of performance ratings is achieved when the number of assessors in OSCE-style assessments is kept limited within stations (1-2 raters) in favour of increasing the number of stations (Eva, Rosenfeld, Reiter, & Norman, 2004; Knorr & Hissbach, 2014).

Future research on the effect of different information sources can turn its attention towards other selection methods such as letters of recommendation or reference letters. These

methods are often used in healthcare selection and calls have been made to improve their reliability and validity (DeZee et al., 2014; Kuncel, Kochevar, & Ones, 2014). Again, a modular approach can be fruitfully used. For example, one might experiment with the type of information source (peers, professors, internship supervisors), the number of referees, or the length of acquaintance with the target person. One could also vary the format of the letters (in the context of varying response evaluation consistency) such as using unstructured, narrative letters vs. letters that require ratings on dimensions that were deemed relevant.

4.7 Effect of Instructions

Instructions are a last building block with key theoretical and practical importance in the design of selection procedures. As with the other building blocks, choices made with this component might affect a selection methods construct saturation, criterion-related validity, or ethnic/gender differences. For example, one can change the amount of specificity or level of detail provided by the response instructions (i.e., instructing participants about what perspective they should take when responding to test items), or switch from knowledge-based instructions to behavioural-tendency instructions ("What should you do?" vs. "What would you do?").

In another study conducted in the Flemish healthcare admission exam, Lievens, Sackett, and Buyse (2009) showed that altering the response instructions from knowledge based ("What should you do?") to behavioural-tendency instructions ("What would you do?) of an SJT had no effects on the scores obtained on the test and did not affect the validity of the SJT for predicting student performance on interpersonally-oriented courses after their first year of healthcare school. They further showed that different response instructions affected the SJT's cognitive loading (i.e., the processing difficulty) only slightly (the relation with cognitive ability test scores was stronger when the response format was knowledge-based). These results led the authors to conclude that in a high stakes context, test takers do not

respond differently when response instructions were changed from one format to the other and that the knowledge-based response format is the preferred choice to minimize "faking good" on the test on the part of the candidates.

Clearly, we need more research on response instructions. An example is the question how much specificity of instructions should be chosen in scenario-based assessment in healthcare selection. Indeed, a dilemma often mentioned is whether one should use general instructions (e.g., How do/did you solve this issue?) vs. more specific instructions (e.g., How do/did you use your interpersonal skills to solve this issue?). With a general instruction one might get insight in how people spontaneously express their trait(s). Situations with little available cues about what to do might also more accurately depict real-life complex decision-making situations. Yet, this low level of instruction specificity might come at a price for the ability to measure constructs of interest since people might fail to give the desired response not because they do not know what to do, but because the situation becomes too ambiguous. This in turn could lead to unreliability in the selection method resulting from its inability to differentiate between people that simply did not know what was expected of them versus people who really did not know the correct answer. Conversely, with more specific instructions it becomes easier to measure constructs of interest but at the same time these instructions could increase the risk of not being able to detect whether people would respond similarly in real life where specific information is also often lacking. At worst, instructions might become so transparent that they give away the correct answer completely, thereby also affecting the ability of a selection method to assess important skills and predict criteria (Ingold, Kleinmann, König, & Melchers, 2016).

5. Case Studies: Putting the Modular Approach to Selection into Action

In this section, we present two case studies. Although the first case study is not situated in healthcare selection, it is a good example of the agility of a modular approach in selection.

Case Study 1: Building New Selection Procedures in Police Officer Selection

In this case, we worked with an organization that had always used a holistic ("all-in-one") approach. The organization had stopped using classic selection centre exercises (because of their less standardized nature and limited coverage of job situations) in favour of a more standardized SJT approach. However, the organization perceived also shortcomings in this SJT approach because actual candidate behaviour was no longer assessed.

This is the point where we decided to adopt a modular approach and move the selection system to the 21st century. In line with the basic axiom of a modular approach, the first thing we did was breaking down these two classic selection procedures (selection centres and SJTs) in their basic building blocks. Presentation format and response format emerged as the two pivotal building blocks. This led to a two-dimensional map as visualized in Figure 1. We then placed the two assessment procedures into this map (see the blue crosses): Selection centres require responding with actual behaviour to the behavioural (role-player) stimuli presented to them, whereas SJTs typically present the situations in a written format and ask people to respond in a written format. As a critical implication, this visualization shows that those two selection procedures that organizations have been using for decades take extreme positions (they are situated at the outskirts of the map). In addition, it becomes clear that there are many empty cells on the map, which might all serve as potential unexplored hybrids between these two extremes. We then decided to create two such “hybrids”, namely (a) an online multimedia simulation with a webcam response format and (b) an online multimedia simulation in which people had to respond by typing their answer in a text balloon.

Using an evidence-based approach, we then set up a field experiment to examine the effects of these two new selection procedures (whose results had no bearing on the selection decision). Results in terms of validity, precision of measurement, and applicant experience

favoured the webcam hybrid. The organization then considered using the modular approach to further adjust this webcam hybrid by modifying another building block, namely response scoring consistency. That is, they chose to score transcribed webcam responses (that had the advantage of not revealing the gender of the candidates) instead of the actual webcam responses.

INSERT FIGURE ONE

Case Study 2: Examining Type of Contextualization in Healthcare Student Admission

In this case study, we worked together with an organization that was involved in designing the SJT of a healthcare student admission procedure. The SJT aimed to assess interpersonal skills. The SJT was regarded as an ideal tool for screening high volumes of candidates to broaden skill coverage (besides cognitively-oriented tests) and increase diversity.

Contextualization is assumed to increase the predictive validity of SJT scores. This assumption is based on the notion that better predictions can be made when there is a point-to-point correspondence between the domain to be predicted (criterion) and the predictor (selection procedures). That is, when SJTs present situations to candidates that closely mirror job-related situations, one can expect that the responses of candidates will be indicative of how they will respond on the job. However, this does not tell us whether the type of contextualization should be framed in “prior familiar realistic situations” vs. “future unfamiliar realistic situations”. A dilemma came up when designing the SJT. One option consisted of embedding the SJT items in a realistic student/educational context. Examples of such a student/educational context are course work, library, and cafeteria. The other option was to frame the items in a realistic healthcare context. In this case, example situations included an ongoing interaction between a physician and a patient. At a practical level, this

dilemma invokes questions about the potential costs and gains of investing in contextualization in SJTs. In case the decision was made to embed the items in a healthcare context, it was clear that (more expensive) subject matter experts would be required for developing (elaborate) situational descriptions, response options, and the scoring key. This was less needed when it would be decided to embed items in a student, educational context.

We framed this dilemma in a modular approach in which we focused on contextualization as one of the key components of selection procedures. We made a distinction between the level of contextualization (low, medium, and high) and the type of contextualization (educational vs. work). This visualization created six cells. In the next step, we decided together with the organization to examine two of these cells. That is, we aimed to compare an SJT with items in an educational, student context to an SJT with items in a work (healthcare) context. The level of contextualization of both of these SJTs was medium. We set up a within-subjects design for examining the effects of type of contextualization. That is, the SJT used in this admissions process consisted of both contextualized items about realistic future healthcare situations (interpersonal physician-patient situations) and contextualized items about realistic past educational situations (interpersonal student-student situations). The effects of this comparison were cast in terms of mean scores (difficulty level), cognitive saturation (correlation with cognitive ability), applicant perceptions, and validity (prediction of future interpersonal course grades). We also scrutinized the potential moderating impact of candidate background (i.e., whether one of the parents had already been working as a health professional) because this might also affect (e.g., potentially bias) the SJT scores.

This examination is currently underway. Results are expected to shed light on which type of contextualization might work best for healthcare admission exams. At a practical level, it might also suggest considerable resource savings in SJT development. So far, lack of

understanding of the effects of contextualization leaves practitioners in the dark when choosing between different types of contextualization.

6. Conclusion

In summary, this chapter presented a modular approach to healthcare selection system design that emphasizes a closer look at the components of selection procedures to complement the traditional holistic view on selection procedures. Our framework identified seven key design components of selection procedures and provided examples of studies in the healthcare selection literature that varied and compared the effect of these components on key selection outcomes. A modular approach has merit in that it allows researchers to gain and provide insights into how the different components underlying selection procedures differentially affect important selection outcomes. Furthermore, such an approach allows drawing conceptual similarities between components of different selection procedures. Lastly, it creates opportunities for experimentation by varying these components and then studying their differential effects on outcomes. Practically, a modular approach permits developing a myriad of new selection procedures by "mixing and matching" different building blocks.

Future research in healthcare selection should apply the modular approach and move beyond typical studies of selection procedures as a whole. Accordingly, such healthcare selection studies should adopt designs where single components are being varied while holding the rest of the selection procedure constant. Although tightly controlled laboratory-style experimental designs with full randomization of participants to conditions are typically viewed as the gold standard for isolating and studying the effects of a variable on important outcomes, the primary and case studies presented in this chapter show that it is also possible to apply the modular approach in quasi-experimental designs with pre-existing samples in field studies. Some shortcomings notwithstanding, such quasi-experimental designs can

provide a good starting point to adopt a modular approach and move the healthcare selection field forward.

Practice Points

- This chapter introduces a modular approach to conceptualizing and examining healthcare selection procedures. This approach departs from viewing selection methods as holistic entities and instead takes a closer look at the "building blocks" or components of selection methods.
- Selection tests can be broken down into seven distinct components. These "Super Seven" are the stimulus format, contextualization, stimulus presentation consistency, the response format, response evaluation consistency, information source, and instructions.
- Each of these building blocks can be isolated and then varied to study their effects (while keeping the rest of the selection method constant) on important selection outcomes.
- (Quasi-)experimental lab and field studies can be set up to conduct research using a modular angle.
- Examining components in isolation reveals their potentially differential effects on important selection outcomes, thereby progressing theoretical and practical understandings of how these components work.
- Drawing conceptual similarities between components across different selection methods also opens up new possibilities to mix and match these components and design new types of "hybrid" selection methods that best suit the clients' goals and needs.

Explore further

Below there is a list of useful articles/books related to a modular approach to selection procedures and a modular approach in general:

Lievens, F., & Sackett, P.R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology, 102*, 43-66. http://doi.org/10.1037/apl0000160

Lievens, F., De Corte, W., & Westerveld, L. (2015). Understanding the building blocks of selection procedures: Effects of response fidelity on performance and validity. *Journal of Management*, *41*, 1604–1627. http://doi.org/10.1177/0149206312463941

Arthur, W. J., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *The Journal of Applied Psychology*, *93*, 435–442. http://doi.org/10.1037/0021-9010.93.2.435

Gershenson, J. K., Prasad, G. J., & Zhang, Y. (2003). Product modularity: Definitions and benefits. *Journal of Engineering Design*. *14*, 295–331. http://doi.org/10.1080/0954482031000091068

Baldwin, C. Y., & Clark, K. B. (2000). *Design Rules: Vol. 1. The Power of Modularity*. Cambridge, MA: MIT Press.

References

Barrows, H. S. (1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Academic Medicine, 68,* 443-451. http://doi.org/10.1097/00001888-199306000-00002

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, *82*, 143–159. http://doi.org/10.1037/0021-9010.82.1.143

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, *63*, 83–117. http://doi.org/10.1111/j.1744-6570.2009.01163.x

Cleland, J., Abe, K., & Rethans, J. (2009). The use of simulated patients in medical education: AMEE Guide No 42. *Medical Teacher*, *31*, 477–486. http://doi.org/10.1080/01421590903002821

De Leng, W. E., Stegers-Jager, K. M., Husbands, A., Dowell, J. S., Born, M. P., & Themmen, A. P. N. (2017). Scoring method of a situational judgment test: Influence on internal consistency reliability, adverse impact and correlation with personality. *Advances in Health Sciences Education*, *22*, 243–265. http://doi.org/10.1007/s10459-016-9720-7

DeZee, K. J., Magee, C. D., Rickards, G., Artino, A. R., Gilliland, W. R., Dong, T., … Durning, S. J. (2014). What aspects of letters of recommendation predict performance in medical school? Findings from one institution. *Academic Medicine*, *89*, 1408–1415. http://doi.org/10.1097/ACM.0000000000000425

Dore, K. L., Reiter, H. I., Kreuger, S., & Norman, G. R. (2017a). CASPer, an online pre-interview screen for personal/professional characteristics: prediction of national

licensure scores. *Advances in Health Sciences Education*, *22*, 327–336. http://doi.org/10.1007/s10459-016-9739-9

Dore, K. L., Reiter, H. I., Kreuger, S., & Norman, G. R. (2017b). Correction to: CASPer, an online pre-interview screen for personal/professional characteristics: prediction of national licensure scores. *Advances in Health Sciences Education*, *22*, 1321–1322. http://doi.org/10.1007/s10459-017-9798-6

Eva, K. W., Rosenfeld, J., Reiter, H. I., & Norman, G. R. (2004). An admissions OSCE: The multiple mini-interview. *Medical Education*, *38*, 314–326. http://doi.org/10.1046/j.1365-2923.2004.01776.x

Harden, R. M., Stevenson, M., Downie, W. W., & Wilson, G. M. (1975). Assessment of clinical competence using objective structured examination. *British Medical Journal*, *1*(5955), 447–51. http://doi.org/10.1136/bmj.1.5955.447

Husbands, A., Rodgerson, M. J., Dowell, J., Patterson, F., Eva, K., Reiter, H., … Nevo, B. (2015). Evaluating the validity of an integrity-based situational judgement test for medical school admissions. *BMC Medical Education*, *15*, 1-9. http://doi.org/10.1186/s12909-015-0424-0

Ingold, P. V., Kleinmann, M., König, C. J., & Melchers, K. G. (2016). Transparency of assessment centers: Lower criterion-related validity but greater opportunity to perform. *Personnel Psychology*, *69*, 467–497. http://doi.org/10.1111/peps.12105

Johns, G. (2006). The essential impact of context on organizational behavior. *Academy of Management Review*, *31*, 386–408. http://doi.org/10.5465/AMR.2006.20208687

Knorr, M., & Hissbach, J. (2014). Multiple mini-interviews: Same concept, different approaches. *Medical Education*, *48*, 1157–1175. http://doi.org/10.1111/medu.12535

Kuncel, N. R., Kochevar, R. J., & Ones, D. S. (2014). A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for hope. *International*

Journal of Selection and Assessment, *22*, 101–107. http://doi.org/10.1111/ijsa.12060

Lievens, F., De Corte, W., & Westerveld, L. (2015). Understanding the building blocks of selection procedures: Effects of response fidelity on performance and validity. *Journal of Management*, *41*, 1604–1627. http://doi.org/10.1177/0149206312463941

Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology*, *91*, 1181–1188. http://doi.org/Doi 10.1037/0021-9010.91.5.1181

Lievens, F., & Sackett, P. R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology*, *102*, 43–66. http://doi.org/10.1037/apl0000160

Lievens, F., Sackett, P. R., & Buyse, T. (2009). The effects of response instructions on situational judgment test performance and validity in a high-stakes context. *The Journal of Applied Psychology*, *94*, 1095–1101. http://doi.org/10.1037/a0014628

Norman, G. R., & Feightner, J. W. (1981). A comparison of behaviour on simulated patients and patient management problems. *Medical Education, 15*, 26–32. http://doi.org/10.1111/j.1365-2923.1981.tb02311.x

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education and training? Evidence from a systematic review. *Medical Education*, *50*, 36–60. http://doi.org/10.1111/medu.12817

Roberts, C., Khanna, P., Rigby, L., Bartle, E., Llewellyn, A., Gustavs, J., … Lynam, J. (2017). Utility of selection methods for specialist medical training: A BEME (best evidence medical education) systematic review: BEME guide no. 45. *Medical Teacher*, pp. 1–17. http://doi.org/10.1080/0142159X.2017.1367375

Swanson, D. B., & van der Vleuten, C. P. M. (2013). Assessment of Clinical Skills With Standardized Patients: State of the Art Revisited. *Teaching and Learning in Medicine*,

25, S17–S25. http://doi.org/10.1080/10401334.2013.842916

Uijtdehaage, S., Doyle, L. H., & Parker, N. (2011). Enhancing the reliability of the multiple mini-interview for selecting prospective healthcare leaders. *Academic Medicine*, *86*, 1032–1039. http://doi.org/10.1097/ACM.0b013e3182223ab7

Van der Vleuten, C. P. M., & D. Swanson, D. (1990). Assessment of clinical skills with standardized patients: State of the art. *Teaching and Learning in Medicine*, *2*, 58–76. http://doi.org/10.1080/10401334.2013.842916

Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology*, *90*, 536–552. http://doi.org/10.1037/0021-9010.90.3.536

Vu, N. V., & Barrows, H. S. (1994). Use of standardized patients in clinical assessments: Recent developments and measurement findings. *Educational Researcher*, *23*, 23–30. http://doi.org/10.3102/0013189X023003023

Waung, M., Hymes, R. W., & Beatty, J. E. (2014). The effects of video and paper resumes on assessments of personality, applied social skills, mental capability, and resume outcomes. *Basic and Applied Social Psychology*, *36*, 238–251. http://doi.org/10.1080/01973533.2014.894477

Weekley, J. A., Hawkes, B., Guenole, N., & Ployhart, R. E. (2015). Low-fidelity simulations. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*, 295–322. http://doi.org/10.1146/annurev-orgpsych-032414-111304

Table 1

Predictor Method Factors, Their Definitions, and Categories (see Lievens & Sackett, 2017).

Predictor Method Factor	Definition	Predictor Method Factor Category/Choice
Stimulus format	Modality by which test stimuli (information, questions, prompts) are presented to test-takers	- Textual stimuli - Pictorial stimuli - Auditory stimuli - Dynamic audio-visual stimuli - Videoconference/remote interactive stimuli - Face-to-face interactive stimuli
Contextualization	The extent to which a detailed context is provided to test-takers	- Decontextualized - Low contextualization - Medium contextualization - High contextualization
Stimulus presentation consistency	Level of standardization adopted in presenting test stimuli to test-takers	- Free stimuli - Adaptive stimuli - Fixed stimuli
Response format	Modality by which test-takers are required to respond to test stimuli	- Close-ended - Textual constructed - Pictorial constructed - Audio constructed - Audio-visual constructed - Videoconference/remote interaction - Face-to-face interaction
Response evaluation consistency	Level of standardization adopted in terms of evaluating test-takers' responses	- Unconstrained judgment - Calibrated judgment - Automated scoring
Information source	Individual responding to the test stimuli	- Behaviour exhibited (or choices made) by the candidate in the assessment context - Self-reports by the candidate about events beyond the assessment context - Reports by others about events outside the assessment context
Instructions	The extent to which directions are made explicit to test-takers about which perspective they should take to respond to the test stimuli	- General instructions - Specific instructions

Figure 1

Predictor Method Map With Crossing of Response Format and Stimulus Format