Learning and Instruction xxx (2017) 1–13

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Instructional sensitivity in vocational education

Viola Deutscher [a, *], Esther Winther [b]

[a] University of Mannheim, Business School, L4 1, 68161 Mannheim, Germany
[b] German Institute for Adult Education (DIE), Heinemannstraße 12-14, 53175 Bonn, Germany

ARTICLE INFO

Article history:
Received 10 November 2016
Received in revised form
30 June 2017
Accepted 7 July 2017
Available online xxx

Keywords:
Instructional sensitivity
Item differential functioning (DIF)
Vocational educational training (VET)
Competence development
Competence-based assessment

ABSTRACT

Apprentices' performance after vocational educational training (VET) is commonly attributed to the effectiveness of the training. This implies the assumption that learners' development of vocational knowledge and ability is significantly affected by vocational instruction. However, the few analyses that have been made of instructional sensitivity within the general school-based educational system, have in most cases shown little or no effect of instruction (time in school) on performance in assessments. The question as to whether, and to what extent, VET in adult education is effective (in the sense that it fosters the development of vocational knowledge and ability), as well as the related question—whether we are able to track the resulting learning progress with adequate measures (i.e., assessments)—has hardly been investigated. In the present study, we propose modeling of instructional sensitivity via differential item functioning (DIF), and apply this method to a sample of n = 534 apprentices. We find that during vocational instruction, apprentices significantly improved their performance in an assessment of vocational knowledge and ability, and that we were able to track these changes in the quality of their abilities over the span of a three year initial VET program: that is, the first program of vocational study in which apprentices become qualified to work in a given trade. Moreover, with this proposed method, it is possible to identify items that are particularly sensitive to instruction and that appear therefore to be amenable to the future development of vocational assessments.

1. Premise

Schooling/training is commonly assumed to be responsible for learning (Burstein, 1989; Naumann, Hochweber, & Hartig, 2014). Somewhat surprising therefore are some empirical hints that performance on assessments in general education is often little or not at all sensitive to the effects of instruction. Diverse research (e.g., Chen, 2012; Court, 2013; Pham, 2009; Phillips & Mehrens, 1988; Popham, 2007; Popham & Ryan, 2012) suggests that many achievement tests fail to effectively reflect whether students successfully receive and absorb curricular content during instruction. This apparent paradox might result from one of two causes (or conceivably both): (1) That learners have indeed learned during instruction, but that the assessment applied was not able to capture the learning progress made. For example, Goe (2007) and Polikoff (2010) caution that the failure to detect instructional sensitivity does not necessarily imply that no learning progress has been made. Rather, the weak relationship between curricular instruction and student performance could be due to the applied measurement tools not being sufficiently sensitive to capture the effect of instruction. These measures of learning outcomes possibly indicate what students know, but not necessarily what they learn during instruction (Popham, 2007).

The second possible cause (2) is expressed by Wiliam (2007, 12) who, providing an insightful analysis of the relevant research addressing instructional sensitivity, goes one step further, arguing for a more pessimistic second order explanation:

> the fundamental issue is not that tests are insensitive to instruction; it is that achievement is insensitive to instruction. Put bluntly, most of what happens in classrooms doesn't change what students know very much, especially when we measure deep, as opposed to surface aspects of a subject.

This second explanation in turn might result from two causes: Either students' knowledge as a latent structure is generally insensitive to instruction, or instruction may not have been delivered (or not effectively).

* Corresponding author.
E-mail address: viola.klotz@bwl.uni-mannheim.de (V. Deutscher).

http://dx.doi.org/10.1016/j.learninstruc.2017.07.004

Please cite this article in press as: Deutscher, V., & Winther, E., Instructional sensitivity in vocational education, *Learning and Instruction* (2017), http://dx.doi.org/10.1016/j.learninstruc.2017.07.004

Even without any clear indication of which of the two explanations (or, conceivably, a combination) accounts for the empirical findings, both interpretations of the instructional *in*sensitivity of diverse outcome measures pose a severe threat—especially to educational accountability. In some nations (e.g., the US), outcome measures have been used in recent times not only to evaluate the effectiveness of schools and teachers on the basis of their students' test proficiency, but also to allocate educational resources on the basis of test results (e.g., state tests used for the purposes of the No Child Left Behind Act). Without a doubt, an accountability test would—as one prerequisite, among other aspects of validity—at least have to be instructionally sensitive, in order to form an appropriate basis for making decisions with potentially far-reaching consequences. However, given unreliable and possibly inaccurate test-based evidence, achievement or learning progress, or even the lack thereof, instructional sensitivity cannot be accurately determined; this leaves the danger that teachers and schools will be misjudged, and even be unfairly denied resources.

Considering these potentially severe consequences, Polikoff (2010, 34), summarizing the overall state of instructional sensitivity research, comes to the conclusion that the lack of documentation of instructional sensitivity in accountability tests constitutes a "grievous oversight". Even more strongly, Popham and Ryan (2012, 2) assail the current lack of empirical evidence regarding instructional sensitivity in most educational tests, describing it as an "intolerable state of affairs". In view of the above, the internationally observable trend towards test-based accountability systems, and political reliance on outcome measures in making decisions affecting education, seems highly questionable. For this reason, some authors have demanded that the concept of instructional sensitivity become an explicit and integral part of a broadened conception of validity, for common standards in educational and psychological testing (e.g., AERA, APA, & NCME, 1999). They call for this to be applied at least for the outcome measures that are used to assess changes in learning and for those testing system effectiveness (e.g., teacher or school effectiveness; for example, Polikoff, 2010; Popham & Ryan, 2012).

Way (2014, 4) raises the concern that "despite these recent imperatives for explicitly making assessments instructionally sensitive, there is not agreement about how this is to be done (…)." Naumann et al. (2014) similarly believe that the question whether outcome measures are indeed sensitive to instruction is hardly empirically engaged, due to the lack of a commonly accepted definition and operationalization of the concept of instructional sensitivity. The methodological approaches to modeling instructional sensitivity are diverse, to say the least: this has led to mainly psychometric papers on the topic, and few practical applications combining the proposed methods with a didactical perspective (for one such application however, see the recent study by Naumann et al., 2014).

Although, as we have noted, instructional sensitivity is a crucial concept in instructional science, to our knowledge no studies have addressed the modeling of instructional sensitivity with respect to vocational education of adults. In Germany, about half of the population takes vocational educational training (VET) rather than academic training, after their school education. Most of this VET (60%) relates to commercial professions: for bankers, industrial management assistants, salesmen (National Educational Report, Hasselhorn et al., 2014). While development of measures of vocational knowledge and ability for this branch of education is very relevant, it is still in its infancy. In general, however, significant progress has been made in the last decade with respect to the measurement of learning outcomes in the vocational domains of auto mechanics (e.g., Nickolaus, Lazar, & Norwig, 2012) and apprenticeships in commercial professions: for example, industrial or logistics apprentices (e.g., Klotz, Winther, & Festner, 2015; Rausch, Seifried, Wuttke, Kögler, & Brandt, 2016; Seeber, 2008; Weber et al., 2016; Winther & Achtenhagen, 2009). More recently, there has also been notable progress in the area of social health care (e.g., Seeber, 2015; Seeber, Ketschau, & Rüter, 2016). Therefore, the purpose of this study is to conceptualize and model instructional sensitivity in the area of vocational education, and to detect which item types are especially relevant to modeling the learning progress. More precisely, we focus on the occupation of industrial management assistant, and seek to explore whether instructional sensitivity is detectable in an assessment of vocational knowledge and ability.

According to Polikoff (2010, 8–9), it is impossible to say

> whether a finding of low or no sensitivity in any particular study is due to a poor-quality test that is actually insensitive to instruction or to poor quality instruction, so that the test results actually reflect the instruction received by students. In contrast, a finding of high sensitivity indicates both effective instruction and also a high-quality assessment sensitive to that instruction. Clearly, the goal is always to have instruction of maximum effectiveness, and to design a test to capture the effects of instruction.

So if we do not find instructional sensitivity, this does not necessarily mean that learners have not learned anything (e.g., due to poor instruction); it may possibly mean that our assessment failed to capture their learning (i.e., instructional *in*sensitivity of the assessment). However, if we find instructionally sensitive items, this must mean that vocational knowledge and ability are being acquired during VET and that we are able to capture them. More precisely, in this study, the following research questions are addressed:

1. Is the developed assessment of vocational knowledge and ability sensitive to instruction (meaning that learning progress is made during VET and that we are able to capture that progress)?
2. Is the learning of specific (vocational) knowledge and ability equally sensitive to instruction as is the learning of generic knowledge and ability?

In order to explore this matter, the paper begins by reviewing different definitions of instructional sensitivity and different methodological approaches to its detection. Subsequently, the item and test design of an instrument to capture apprentices' knowledge and ability is introduced. We then apply the IRT-DIF approach to a vocational sample of n = 877 industrial apprentices, and outline and discuss the results.

2. Defining and detecting instructional sensitivity

In the theoretical research into instructional sensitivity, this term has often been used interchangeably with "instructional validity", with both terms being treated as subfacets of other, common aspects of test validity, such as curricular validity and content validity (Polikoff, 2010). Li et al. (2012b, p. 2) note that the intended meaning of the term sometimes relates exclusively to the extent to which the curriculum content is taught successfully (e.g., Linn, 1983). Occasionally however, it also includes the nature of the teaching of the content (e.g., Burstein, Aschbacher, Chen, & Lin, 1990; Popham & Ryan, 2012; Yoon & Resnick, 1998). A definition that is open to both interpretations is the originally used, more technical definition of Haladyna and Roid (1981, p. 40), defining instructional sensitivity as "the tendency for an item to vary in difficulty as a function of instruction". This relation is then specified

Please cite this article in press as: Deutscher, V., & Winther, E., Instructional sensitivity in vocational education, *Learning and Instruction* (2017), http://dx.doi.org/10.1016/j.learninstruc.2017.07.004