

Interviewer-Respondent Interaction and Rapport in PIAAC

Bryan Maddox

Abstract

Purpose: The paper investigates small-scale, qualitative observations of interviewer-respondent interaction in the Organisation for Economic Co-operation and Development's Programme for the International Assessment of Adult Competencies (PIAAC).

Design/methodology/approach: The paper employs video-ethnographic methods to document talk and gesture in assessment in Slovenian household settings. It presents an in-depth case study of interaction in a single testing situation.

Findings: Observing interaction in assessment captures data on assessment performance that is not available in quantitative analysis of assessment response processes. The character of interviewer-respondent interaction and rapport is shaped by the cognitive demands of assessment and the distinctive ecological setting of the household.

Research limitations/implications: Observational data on assessment response processes and interaction in real-life assessments can be integrated into and synthesized with other sources of "process data."

Practical implications: Assessment programs such as PIAAC should consider the significance of the household setting on assessment quality and observations of interaction in assessment as a valid source of paradata.

Social implications: There is a place for small-scale observational studies of assessment to inform public understanding of assessment quality and validity.

Originality/value: The paper provides qualitative insights into the significance of interaction and “interviewer effects” in household assessment settings.

Keywords: households, assessment, interaction, rapport, interviewer effects

Introduction

This paper examines the significance of interviewer-respondent interaction in the Organisation for Economic Co-operation and Development’s (OECD’s) Programme for the International Assessment of Adult Competencies (PIAAC) and its relationship to data quality. It presents a qualitative case study that questions the role of this interaction on data quality, and considers its significance in the context of household-based assessment. The literature on Conversation Analysis promotes the view that interaction in the form of talk and gesture in face-to-face social situations is integral to the way that social situations are produced and performed (Goodwin and Duranti, 1992). The importance of interaction has not gone unnoticed by researchers in educational assessment, who note the significance of talk in producing interviewer and respondent roles and relations, and its implications for assessment performance and consequences (McNamara, 1997; Shohamy, 1998; McNamara and Roever, 2006; Maddox, 2015). This raises questions about the merit of interaction in assessment, why it occurs, and how it is treated in interviewer training and in quality assurance practices.

PIAAC is a large-scale assessment of the adult population, aimed at collecting comparable information on literacy, numeracy, problem-solving ability in technology-rich environments as well as background information on around 200,000 adults worldwide (5,000

respondents in each country). The assessment normally takes place in the respondent's home. It is either delivered by computer (the focus of this paper) or in a paper-based mode for respondents who are not sufficiently able to use the computer. The target population for PIAAC is the noninstitutionalized adult population, ages 16-65, residing in the country at the time of data collection, irrespective of nationality, citizenship, or language status. The survey is administered in the official language or languages of each participating country; some countries gave respondents the possibility of participating in one of the widely spoken minority/regional languages. The first round of PIAAC (2008-2013) involved 24 countries/national subregions. A second round (2012-2016) involved 9 countries including Slovenia, which is the focus of this paper. A final round in the first cycle takes place from 2016-2019 and includes a further 5 countries. ^[i].

In the recent literature on standardized interviewing and household surveys there has been some useful debate about the merit of interviewer improvisations as departures from the standardized script. From one perspective, high-quality data is associated with standardization, and departures from the script are portrayed as a source of negative “interviewer effects” that introduces unwelcome sources of variation, bias, and error. Such improvisations are associated with interviewer identity, attitudes, and personality (see Ackerman-Piek and Massing, 2014; Bell *et al.*, 2016; Brunton-Smith *et al.*, 2017). A contrasting perspective considers the potential for content and task appropriate interaction to impact positively on respondent participation, engagement, and motivation (Garbarski *et al.*, 2016). From that perspective, interviewer departures from the script can involve content and task appropriate interaction and rapport—such as clarification about items, displays of empathy, and encouragement (Garbarski *et al.*, 2016, p. 3). As they argue: “Rather than viewing rapport as a violation of standardization, we examine whether the behaviors that constitute responsiveness and engagement complement or conflict with the practices of

standardization to accomplish the task of obtaining codeable answers to survey questions.”

This paper adopts a similarly agnostic stance on the status and contribution of interviewer-respondent interaction in PIAAC.

While it recognizes the potentially beneficial and necessary role of improvised interaction, this paper attempts to push the debate further by asking why such interaction should be necessary in computer-administered assessment. That is, it examines the purpose and significance of the interaction, and what it can tell us about the quality of test administration.

PIAAC can be distinguished from a conventional survey because the assessment component presents the respondent with challenging cognitive tasks, and a necessary feature of assessment design is that the respondent should *not* be able to complete all the tasks. As a result, the tasks are likely to push the respondent to the limits of their resilience. Furthermore, the “low stakes” survey design of PIAAC is not like a conventional test, as the respondents are selected on a randomized basis and should be unprepared for the assessment (i.e., they do not revise or prepare in advance). Even with its multistage adaptive design, the assessment is particularly challenging for respondents who are at the lower end of the ability spectrum and achieve correct answers on the least number of assessment tasks. These challenging characteristics of the PIAAC assessment frequently prompt interviewers to provide assistance and to demonstrate rapport, empathy, and encouragement.

In the computer-administered version of PIAAC, the “script” (the delivery of instructions and test items) is largely delivered by the computer. After the completion of the background questionnaire (BQ), the interviewer hands over the laptop to the respondent, who then follows the instructions and completes the test items that appear on the screen. In an ideal situation, the computer would provide all the information and support that the

respondent requires to complete the assessment. However, there are some areas where the human interviewer is either *more able* to provide support (e.g., such as emotional support and encouragement), or where respondents find the interviewer to be a *preferable* source of support (e.g., such as on troubleshooting the correct procedures for item completion, or providing information about time required to complete the assessment). Often, the mere presence of the human interviewer and his or her verbal and gestural displays of empathy invite the respondent to look to the interviewer rather than the computer as a collaborator in the testing situation. As the case study illustrates, interviewer-respondent interaction in PIAAC often has a complementary or compensatory role—involving information and support that is not adequately provided by the computer, that is, as a case of human-computer-human interaction.

Households and Assessment

PIAAC is a household-administered survey and assessment. As we will see below, this creates some ecological dynamics that impact on the content and demands of interviewer-respondent interaction. The implications of this ecological setting merit careful consideration.

Garbarski *et al.* (2017, p. 2) suggest that interaction in survey interviews should be appropriate to their “specific task environment.” We can therefore consider how the household setting might influence appropriate and quality-oriented interaction in large-scale, standardized assessment.

The household is not necessarily a good proxy for a testing center. Taking the test to the respondent is one way to encourage higher response rates in assessment. However, the household setting can place unusual demands and obligations on interviewers that make

quality assessment difficult. Households are inherently “noisy” ecological settings. Telephones ring and televisions play. Assessment that takes place in household settings is prone to interruption and disturbance by family members and friends—particularly the presence of children as “bystanders” (Goffman, 1981) who may want to participate in or influence the outcome of the assessment. As Goffman (1981, p. 132) suggests, the presence of such participants in face-to-face situations may be considered “the rule, not the exception,” and this observation probably holds true for PIAAC assessment settings.

The everyday routines of the household—such as cooking and preparing meals and the requirements of child care—may have an impact by placing demands on the respondent and the time that they have available to participate, and those demands are likely to have age- and gender-based characteristics. Furthermore, households are not singular or self-evident institutions (Randall *et al.*, 2011; Yanagisako, 2015). The choice of a household setting for assessment introduces a potential source of construct-irrelevant variation within and across countries, as there are significant differences in the family structure, social practices, and meanings attributed to households.

As a physical and material space, there is significant variation within and across cultures as to what a household involves (Randall *et al.*, 2011). Its design and cultural features influence how much privacy an individual respondent can receive, or indeed, on how much privacy it is *appropriate* for an interviewer and respondent to be given when conducting an assessment. In some households it is an obligation for family members to offer “help” to the respondent. It may also be appropriate or necessary in some settings for respondents to complete the assessment while attending to child-care responsibilities, while in others the respondent is offered privacy and quiet. Those kinds of differences and the types of interaction that they imply are evident for example, if we contrast assessment conducted in

single-room *gers* (the traditional, round, felt-covered nomadic homes) observed in the Mongolian LAMP (Literacy Assessment and Monitoring Programme) assessment (Maddox, 2014, 2015) with the spacious family homes and kitchen-diner location of assessment observed in the Slovenian PIAAC assessment (discussed below).

Those characteristics of the household shape the kind of interaction that takes place in assessment, and the kinds of interviewer performance that we might consider characteristic of good assessment.

Method

The data presented in the paper are the product of video-ethnographic observations conducted during the Slovenian PIAAC assessment in 2014. The observations contribute fine-grained “process data” (see Ercikan and Pellegrino, 2017; Zumbo and Hubley, 2017) that can be used to understand the character of interviewer-respondent interaction in PIAAC.

Video-ethnographic methods were used to observe a series of PIAAC testing situations (n=12) involving 5 male and 7 female respondents. The research involved noninvasive observations of real-life assessment practices as the ethnographer accompanied interviewers in the course of their normal work. The testing situations took place in rural and peri-urban contexts across several geographical areas of Slovenia.

The method that informs this paper is microethnographic as its concern is with face-to-face interaction (talk and gesture) in real-life testing situations rather than the wider social and institutional context of the assessment (see Maddox, 2015; Maddox, 2017). The approach can be described as inductive since it seeks to observe social phenomenon, and to identify patterns that can inform explanation and the development of hypotheses. The approach

contrasts with the deductive orientation of “process model interpretations” described by Kane and Mislevy (2017).

The paper contributes to a wider literature that examines the content of talk to investigate the quality and consequences of interaction in standardized surveys (Ackerman-Piek and Massing, 2014; Garbarski *et al.*, 2016). It is informed theoretically and methodologically by Conversation Analysis and Gesture Studies (Duranti and Goodwin, 1992; Goodwin, 2007), with attention to sequences of action-oriented talk and gesture as interviewers and respondents complete the assessment (see also Maddox and Zumbo, 2017). The focal feature of the paper is therefore on the goal-oriented “ecological huddle” (Goffman, 1964, 135) connecting the interviewer, respondent, and computer.

Observing testing situations produces detailed process data on real-life assessment practice (Maddox and Zumbo, 2017, Maddox, 2017). This includes ethnographic descriptions of talk and gesture, located in the wider context of the testing situation. These can be considered as “in vivo” observations (Maddox and Zumbo, 2017) in the household setting, in contrast to the more contextually sterile “in vitro” setting of laboratory-based observations. In this way, the approach deliberately incorporates ecological dimensions into the analysis of assessment performance, that is, as potentially telling characteristics, even if they are not considered construct relevant (Zumbo, 2007).

In this study, off-camera observations of household interactions were also recorded as ethnographic field notes. Those wider observations help to explain how the social and material space of the household impacts on assessment response processes. “Being there” to observe real-time behavior in a testing situation enables various sources of information on test taking behavior to be interpreted together as a holistic interpretation of response processes. That includes making sense of microanalytic data about interaction, such as

interpretation of subtle changes in intonation (prosody), the significance of a sigh, a shake of the head, or a blushing face.

Interviewers in PIAAC must be skilled in managing respondent questions and interventions. They must also be attentive to the emotional stance of the respondent and, where necessary, be prepared to encourage the individual to continue to the end of the test (Organisation for Economic Co-operation and Development, 2011). Skilled interviewers will respond sensitively to the respondent's stance as it is expressed in facial expression and body posture. They will be able to respond to questions and to potential threats to the integrity of the testing situation from other family members. Those roles are anticipated in the interviewer manual—that is, interviewer improvisations are part of the normal interviewer training and protocols for managing the testing situation (Organisation for Economic Co-operation and Development, 2011).

Observing the fine detail of assessment can help distinguish between a straightforward reaction to the cognitive demands of a test item and the responses to those wider ecological settings that can impact on assessment performance. This can typically include responses to inquiries from children and other family members who may be present and entail such responsibilities as child care, demands for play, and food preparation (i.e., typical household activities). The demands and responsibilities of household members can impact on how time is perceived and experienced and influence “fatigue” responses such as disengagement and item skipping.

The example below is chosen as a telling case study of interaction in PIAAC. It is not representative of testing practice but reveals, in a single example, some of the characteristics of interaction present in testing situations.

Observations

'I really don't feel like calculating right now.'

One of the distinguishing features of an international large-scale assessment such as PIAAC is that the participants are not fully prepared for the assessment, and may not be entirely willing participants. They have not prepared in advance, and usually have little prior knowledge about, what the assessment will involve.

One of the roles of the interviewer is to persuade respondents to agree to participate in the test, and if they agree, to maintain their active engagement and effort for the duration of the test. Those efforts of persuasion and encouragement may be especially pressing for respondents with low levels of ability. For these respondents, literacy and numeracy tasks may be associated with high levels of anxiety. The following case study concerns one of those examples. The respondent, a woman in her 40s with three children, was initially reluctant to participate. The interviewer spent at least 15 minutes discussing the assessment with the respondent before she agreed to participate and invite us into her house.

Like the other PIAAC assessments that I observed in Slovenia, the completion of the BQ took place in the kitchen-dining room. I sat with the interviewer, and the respondent answered questions while she tidied the room and washed some dishes. The respondent chatted freely during this process, and this enabled the interviewer to establish rapport. But as a result of this discussion, the BQ process took longer than it otherwise might have. The interview began at 3:45 p.m. and the assessment was not complete until 6:15—at 2½ hours, considerably longer than most assessments.

As she started the computer-based assessment, the respondent sat at the kitchen table with the interviewer sitting next to her. I sat at the other side of the table and did not converse

with the interviewer or respondent during the BQ or the assessment (helped by my inability to converse in Slovenian). While the BQ took place, the respondent's children were playing with a football outside. The initial communication between the interviewer and respondent (37 seconds into the assessment) concerned procedures for answering the items. In almost all of the assessments observed, respondents asked procedural questions within the first couple of minutes.

R: Where should I write it?

I: I'm not allowed to help you to solve the exercises, so, please, solve them independently. ... You move on to the next exercise by clicking 'next.'

R: Haven't I, I have clicked... I see! Hmm, just a moment ...

This short exchange illustrates two aspects of interviewer-respondent interaction in PIAAC. The respondent views the interviewer as a source of information and assistance, and the interviewer uses the occasion to set out some ground rules such as, "I am not allowed to help you to solve the exercises." That exchange is typical of many of the testing sessions observed.

As she gets into the assessment, the respondent communicated her feeling of unhappiness through nonverbal "response cries" (Goffman, 1981) such as facial expressions and by gently shaking her head. Her face was a bit flushed. The respondent talked quietly to herself (no longer to the interviewer) as she completed the tasks. Then, at 7½ minutes, the respondent expressed her feelings more clearly to the interviewer:

R: I really don't feel like calculating right now! ... Can I go on if I don't feel like making the calculation?

I: The decision is yours.

R: My brains are not functioning right now. This is 12 ... ah, I'll move on!

I: If you need a few minutes break, take it.

R: No, you know what, I didn't have a clue it would be something like this!

I: You now also have an opportunity to learn something new. Don't be afraid of anything. These are now the instructions for the short exercises, would you read them, please?

R: Hmm, what is this ... do I click "next"?

I: You always move forward on this part of the monitor.

In the above sequence, the respondent expresses her stance or orientation to the tasks in the numeracy module: *I really don't feel like calculating right now*. She also asks a procedural question: *Can I go on if I don't feel like making the calculation?*, and receives a response from the interviewer that might have been drawn straight from the interviewer manual—*The decision is yours*. The short exchange contains some important information as the respondent provides feedback that the test was not what she had expected. In response, the interviewer demonstrates rapport to provide some moral support and encouragement.

The respondent continues with the assessment, with occasional interruptions or advice about procedures for submitting an answer (at 10 minutes), and about the operation of the mouse (at 15 minutes). Then, at 21 minutes into the assessment, things get a bit awkward. As she had been completing the items, the respondent had been communicating her unhappiness with a series of sighs and shakes of her head. Then as she looked directly at the

interviewer (not at the screen), the respondent asks if someone else might do the assessment for her.

R: Could I pass this to ...

I: Pardon?

R: Could I pass this to someone who can do it faster?!

I: Look, take as much time as (you need).

R: Yeah,.. you know I have time limits! Time limits! This (pointing to the screen) the second yes, 6½ times. Can you help me because I don't know how to correct this? ...

Is it necessary that I do this?

I: Please. If you can, if you're so kind.

This exchange demonstrates an effort by the respondent to obtain help from the interviewer. The interviewer gently refuses to provide help (with a smile and a shake of her head) and the respondent continues with the assessment. However, a few minutes later, at 24 minutes into the assessment the respondent expressed her feelings more strongly:

R: It won't work. I don't have any will, not to read or anything else.

I: Take a few minutes' break.

R: No, you know, because someone ... someone is spurring me. The evening is almost over and I promised my child we would take a walk.

I: I see, you told her you would go out?

R: Yeah, I promised her, already yesterday. I have been absent a whole day, and I told her that we would go today. And you're pestering me with this ... I thought that I would be merely answering, not that it would be so tiresome.

I: There are only (a) few short exercises left.

R: No, I don't have any will left. When my morale is gone, it's gone totally.

I: Why?

R: I don't know.

I: Do you have a bad feeling?

R: I have a bad feeling and [I]

I: Don't feel bad about it.

R: Ah, I don't feel like doing it at all!

I: Look at the next short exercises ...

R: When I don't feel like ... at all ... my mind doesn't ... work, really. I had completely different plans for today, really ...

I: You should have told me ..

R: Yeah, how could I know? I thought that you would only ask me questions and there would be enough time for everything!

I: We really don't have many exercises left. Some, would you just go through them? Could you do some more, please?

The exchange highlights themes of motivation discussed by Eklöf (2010, p. 345) in which she argues that assessment performance is a “joint function of skill and will.” The sequence above appears to confirm Eklöf’s observations. In this case it suggests that respondent stance and motivation are not simply static characteristics but, rather, dynamic and contingent upon the ecological setting of the household and on interaction with the interviewer. The respondent claims that she does not have any will left, and the interviewer’s improvised response encourages her to continue.

After that exchange the respondent continued to complete the assessment. However, in the process, the dynamics of the household-based testing situation became further complicated by the presence of additional family members. Three children arrived on the scene, initially in search of biscuits, but on observing the testing situation, they decided to join and watch their mother complete the assessment. At 35 minutes into the assessment there are 6 people present – the respondent, her 3 children, the interviewer, and the ethnographic observer. The presence of the children certainly lightened the mood but also introduced new challenges for the interviewer as the children inquired about the assessment and offered their suggestions and help.

R: Look, you have three times more time than I have! Will you solve it instead of me?!

Child 1: What?

I: They can’t solve it instead of you. They will do it on some other occasion.

R: If you do the washing instead of me.

CI: No.

R: You will solve this instead of me, you have much more time.

C1: What is a poll?

C1: I can't read (can't see the screen) ...

I: You can't read?

C1: No.

I: Slovenian or English? (smiles, joking)

C1 : No, I know how to, I know.

I: I know you know (smiles).

R: Help me here!

C1: What?

I: You can't help her, because ... we asked her if she could do it.

C2: But you only have to click, and that's it!

R: So come and click instead of me! Do you think I have as much (time as you do)?

I: But you're doing all right.

The interaction shows how the children provide moral support to their mother and the interviewer ensures that they don't actually provide the answers. At that point the children ask about the purpose of the assessment and how the computer works. The interviewer is able to explain. That prevents them from intervening directly in the assessment process, and encourages their mother to continue and complete the assessment.

I: Your mummy is in school now (smiles)

C1: Which one do you expect ...

R: Can you leave her to solve it on her own?

C1: I won't help her any more. Even if I could.

I: You know well ... You're doing really fine.

R: I'm not doing fine at all, my head is completely elsewhere.

C1: Well..! (joking)

I: Believe me, you're doing really fine.

C1: You should only concentrate.

R: Yeah, you can say that to me!

C1: Yeah, I can, yes!

R: Yeah, because you have time, and I don't have it.

C1: I don't have time.

R: And if I ask you to do something instead of me, you won't do it!

C1: I can't do this.

C2: Yeah!

R: Well, see, that's how it is!

C1: Well, yeah!

R: Yeah! I'll pay you back, wait and see!

As we can see in the sequence above, the respondent continues to frame the demands of the tasks in terms of time available. The impression is that her earlier reluctance to complete the assessment was a response to the challenging cognitive demands of the assessment, and the wider demands of household responsibilities. The theme of *time* is raised in the discussions among the interviewer, respondent and her family. The following interaction takes place 49 minutes into the assessment:

R: What have we got here, ah ... I wish I could have half an hour for myself!

I: Here you have an opportunity to learn something new. These short exercises are in a way expressions or imitations of real-life situations. All the information you see here are also useful. They are collected from certain ... several sources.

CI: How many exercises?

I: Pardon?

CI: How many exercises are there?

I: That depends solely on an individual. The exercises are determined from the corpus of the designed exercises, so not everybody is solving the same exercises.

I: No two people are solving the same exercises?

CI: It depends on how you react probably.

I: And on other things ... I don't know what algorithm is behind the selection of exercises. It depends, as first, on the individual, to some extent also on one's previous answers and manner of solving the exercises. It is adjusted to individuals and the situation.

CI: How much time does it take on average?

I: On average ... This again depends on the population I work with. ... It also depends on how much you talk with participants while surveying the questionnaire. Some don't ask question at all, don't ask for explanations, so it is all much faster.

R: Good, now I will stop asking questions! (referring to C1)

I: It all really depends on an individual.

Here the child has become the advocate for their mother. She raises the question of time, and the child continues the discussion. The interaction highlights one of the challenges of PIAAC—that the time for the tasks can vary considerably based on the ability of the respondent and the allocation of tasks by the computer (based on the multistage computer-adaptive test; CAT). In response, the interviewer becomes the advocate for the computer and its algorithm. The question of how long the assessment will take was frequently raised during the assessments observed in Slovenia.

Unlike computer games that provide a constant supply of encouragement and information about how long the “player” has to go, the computer in the PIAAC assessment does not provide encouragement, or tell the respondent how far they are through the assessment process. There are other occasions where the interviewer must step in to “help” the computer. This includes any software or hardware glitches that might occur. In the case of this assessment, the respondent has had difficulty operating the mouse, and her hand and arm have started to ache. For example, at 26 minutes into the assessment, the following interaction takes place.

R: Why doesn't it work? (referring to the mouse)

I: Pardon?

R: Why doesn't it work?

I: Try again.

Then at 58 minutes into the assessment it appears that physical fatigue has become an issue as well as mental tiredness:

R: My hand aches.

I: If your hand aches, do this: Massage it a bit, where it hurts, and it will go away.

Then at one hour into the assessment:

R: I see it, my arm aches. Because I constantly have to ...

I: Where exactly does it hurt?

R: The entire arm. Especially now, after sitting for a longer while. And up here too, all (rubbing her arm) ...

I: Move over to this side a bit ...

R: No, it won't help. It starts to hurt here inside ... and then all ...

I: In such a case, you do this (suggesting how to relax her arm). Only this. So you relax the muscle a bit.

R: I see. Oh, tell me when do we end?

Discussion

The literature on interviewer effects and rapport in standardized surveys debates whether “off script” improvisations are threats to data quality or necessary and appropriate

responses to item and task content (e.g., Ackerman-Piek and Massing, 2014). This paper contributes to that literature by examining the content of interviewer-respondent interaction in PIAAC computer-administered testing situations.

The methodological contribution of the paper has been to demonstrate that small-scale, video-ethnographic observations can provide fine-grained “process data” on interviewer-respondent interaction in real-life testing situations. That data includes information on talk, facial expression, and gesture that is not available from other sources of process data such as computer-generated log-files.

This “bottom up” inductive research has used observations of assessment practice to identify three hypotheses that help to explain the presence and significance of interviewer-respondent interaction and rapport in PIAAC.

The first is that the PIAAC assessment creates challenging cognitive demands on the respondents who are at the lower end of the ability spectrum. As a result, interviewers provide information and encouragement, and signal empathy (i.e., as qualities of rapport) to maintain the respondent’s motivation and engagement. As the case study illustrates, the respondent may also ask for help, and a skilled interviewer should be able to resist those requests.

Second, the video-ethnographic observations suggest that the computer-administered test does not provide all the information and encouragement that some of the PIAAC respondents require. Information on timing and progress is largely absent in the CAT design, and its absence is one source of interviewer-respondent interaction. Similarly, in a marked difference to gaming software, the PIAAC computer does not encourage respondents, and as

a result, encouragement is provided by the interviewer, who can perceive and respond to the respondent's affective state.

Third, the fact that the PIAAC assessment is conducted in the household involves numerous threats to the integrity of the assessment. The household is a “noisy” ecological space, and one that varies within and across cultures. Threats to the integrity of the testing situation include interventions from bystanders who are not ratified participants (Goffman, 1981)—such as family members or friends who may inquire about the test, offer support to the respondent, or volunteer to take the test themselves.

To conclude, we can consider the significance of these findings for the next cycle of PIAAC and for similar types of household-based assessments. As this research was inductive and exploratory, its purpose was to identify and explain phenomena. It involved a small opportunistic sample in one country. This suggests that some caution is necessary about making recommendations for future test administration. However, as an example of “process scrutiny”, (Newton, 2016), the research identified some themes that deserve further attention.

The observations described in this paper confirm assumptions that are implicit in the PIAAC interviewer manual—that interviewer interventions and rapport in the assessment process are a potential source of variation in data quality. Further work would be necessary to investigate whether those observations are representative of the experience with PIAAC in other countries and to distinguish between behaviors that represent good quality test administration from those that are a source of unwelcome interviewer effects. That can be achieved through further ethnographic observations, and through investigation of larger-scale sources of data on assessment process and performance.

Finally, the observations relating to human-computer interaction merit investigation and further work on design for the next cycle of PIAAC. Improved information from the computer on timing or progress through the assessment, and additional encouragement from the computer, could reduce the burden on the interviewer and help standardize assessment administration. Such improvements are unlikely to be able to anticipate all the ecological challenges involved in household-based assessment. As a result, interviewer-respondent interaction looks certain to play a continued and important role in the performance of quality assessment.

References

- Ackerman-Piek, D. and Massing, N. (2014), "Interviewer behaviour and interviewer characteristics in PIAAC Germany", *Methods, Data, Analyses*, Vol. 8 No. 2, pp. 199-222.
- Bell, F., Fahmy, E. and Gordon, D. (2016), "Quantitative conversations: the importance of developing rapport in standardised interviewing", *Quality and Quantity*, Vol. 50, pp. 193-212.
- Brunton-Smith, I., Sturgis, P. and Leckie, G. (2017), "Detecting and understanding interviewer effects on survey data by using a cross-classified mixed effects location-scale model", *Journal of the Royal Statistical Society*, Vol. 182 Part 2, pp. 551-568.
- Duranti, A. & Goodwin, C. (Eds.). (1992), *Rethinking context: language as an interactive phenomenon*. Cambridge University Press, Cambridge, UK.
- Eklöf, H. (2010), "Skill and will: test-taking motivation and assessment quality", *Assessment in Education: Principles, Policy and Practice*, Vol. 17, pp. 345-356.

- Ercikan, K. and Pellegrino, J.W. (Eds.) (2017), *Validation of Score Meaning for the Next Generation of Assessments*. Routledge, Abingdon, UK.
- Garbarski, D., Schaeffer, N.C. and Dykema, J. (2016), “Interviewing practices, conversational practices and rapport: responsiveness and engagement in the standardized survey interview”, *Sociological Methodology*, Vol. 46 No. 1, pp. 1-38.
- Goffman, E. (1964), “The neglected situation”, *American Anthropologist*, Vol. 66, pp. 133-136.
- Goffman, E. (1981), *Forms of Talk*. University of Pennsylvania Press, Philadelphia, PA.
- Goodwin, C. (2007), “Environmentally coupled gestures”, in Duncan, S., Cassell, J. and Levy, E. (Eds.), *Gesture and the Dynamic Dimension of Language*, John Benjamin’s Publishing, Amsterdam and Philadelphia, pp. 195-212.
- Goodwin, C., and Duranti, A. (1992), Rethinking context: an introduction. In A. Duranti, and C. Goodwin (Eds.), *Rethinking context: language as an interactive phenomenon* (pp. 1–42), Cambridge University Press, Cambridge, UK.
- Kane, M. and Mislevy, R. (2017), “Validating score interpretations based on response processes”, in Ercikan, K., and Pellegrino, J.W. (Eds.), *Validation of Score Meaning for the Next Generation of Assessments*. Routledge, Abingdon, UK, pp 11-24.
- Maddox, B. (2014), “Globalising assessment: an ethnography of literacy assessment, camels and fast food in the Mongolian Gobi”, *Comparative Education*, Vol. 50, pp. 474-489.
- Maddox, B. (2015), “The neglected situation: assessment performance and interaction in context”, *Assessment in Education*, Vol. 22 No. 4, pp. 427-443.

- Maddox, B. (2017), "Talk and gesture as process data", in *Measurement: Interdisciplinary Research and Perspectives*, Issue 3-4, pp. 13-127.
- Maddox, B. and Zumbo, B.D. (2017), "Observing testing situations: validation as jazz", in Zumbo, B.D., and Hubley, A.M. (Eds.), *Understanding and Investigating Response Processes in Validation Research*, Springer Press.
- McNamara, T. (1997), "Interaction in second language performance assessment: whose performance?" *Applied Linguistics*, Vol. 18 No. 4, pp. 446-466.
- McNamara, T. and Roever, C. (2006), *Language Testing: The Social Dimension*. Blackwell, Oxford.
- Newton, P. (2016), "Macro- and micro validation: beyond the 'five sources' framework for classifying validation evidence and analysis", *Practical Assessment, Research and Evaluation*, Vol. 21 No. 12, <http://pareonline.net/getvn.asp?v=21&n=12>
- Organisation for Economic Co-operation and Development (2011), *Interview Procedures Manual. PIAAC Main Study*, OECD, Paris.
- Organisation for Economic Co-operation and Development (2016), *Technical report of the Survey of Adult Skills (PIAAC)*, OECD, Paris.
- Randall, S., Coast, E. and Leone, T. (2011), "Cultural constructions of the concept of household in sample surveys", *Population Studies*, Vol. 65 No. 2, pp. 217-229.
- Shohamy, E. (1998), "Critical language testing and beyond", *Studies in Educational Evaluation*, Vol. 24 No. 4, pp. 311-345.

Yanagisako, S.J. (2015), “Households in anthropology”, in *International Encyclopedia of the Social and Behavioural Sciences*, 2nd Edition, Vol. 11, pp. 228-232.

Zumbo, B.D. (2007), “Three generations of DIF analysis: considering where it has been, where it is now, and where it is going”, *Language Assessment Quarterly*, Vol. 4, pp. 223-233.

Zumbo, B.D. and Hubley, A.M. (Eds.) (2017), *Understanding and Investigating Response Processes in Validation Research*. Springer.

^[i] See OECD (2016) and www.oecd.org/site/piaac for technical details.