

Jan Nehyba and
James Lawley

Clean Language Interviewing as a Second-Person Method in the Science of Consciousness

Abstract: *This article reports on Clean Language Interviewing (CLI), a rigorous, recently developed ‘content-empty’ (non-leading) approach to second-person interviewing in the science of consciousness. Also presented is a new systematic third-person method of validation that evaluates the questions and other verbal interventions by the interviewer to produce an adherence-to-method or ‘cleanness’ rating. A review of 19 interviews from five research studies provides a benchmark for interviewers seeking to minimize leading questions. The inter-rater reliability analysis demonstrates substantial agreement among raters with an average intraclass correlation coefficient of 0.72 (95% CI). We propose that this method of validation is applicable not only to CLI but to second-person interviews more generally.*

Keywords: second-person interview; Clean Language Interviewing; evaluation; qualitative validation; content-empty perspective.

1. Introduction

The elicitation of first-person descriptions by second-person methods in the science of consciousness has become increasingly sophisticated. Second-person phenomenological methods that explore human

Correspondence:
Email: jan.nehyba@gmail.com

consciousness and experience involve a range of ‘interview techniques that solicit both verbal and non-verbal information from participants in order to obtain systematic and detailed subjective reports’ (Olivares *et al.*, 2015, p. 1). Varela’s (1996) ideas, for example, were developed into a specific version of the ‘elicitation interview’ (Petitmengin and Bitbol, 2009; Vermersch, 1999), now called the ‘micro-phenomenological interview’ (Bitbol and Petitmengin, 2017). Other methods of eliciting inner experiences include the expositional interview which uses Descriptive Experience Sampling (DES; Hurlburt, 2011a) and lesser known approaches involving hypnosis (Lifshitz, Cusumano and Raz, 2013) and neuro-linguistic programming (NLP; Tosey and Mathison, 2009). Gendlin’s (2004) ‘Thinking at the Edge’ method makes an important contribution to the discussion about how to obtain first-person data with the support of second-person interviews. Furthermore, methods of validation and evaluation of second-person interviews have been the subject of debate in the *Journal of Consciousness Studies* (Froese, Gould and Seth 2011; Hurlburt, 2011b; Petitmengin and Bitbol, 2011).

This article aims to extend the range of second-person interview methods by reporting on a rigorous, recently developed approach called Clean Language Interviewing (CLI). It also aims to add to the validation debate by presenting a new systematic third-person method that evaluates interviewer questions and statements for (a) the ratio of leading/clean interviewer interventions, and (b) adherence to interview design and protocols.

2. Content-Empty Perspectives

A feature of second-person interviews is the specific way of formulating questions, variously referred to as non-leading (Depraz, Varela and Vermersch, 2003), content-empty (Bitbol and Petitmengin, 2017; Petitmengin, 2006), open-beginninged (Hurlburt and Schwitzgebel, 2007), and clean questions (Tosey, Lawley and Mees, 2014). These questions encourage interviewers to minimize assumptions that could influence interviewees’ descriptions of their experience; in other words, to ‘cleanse our phenomenological palate’ (Hurlburt and Heavey, 2006, p. 86).

Olivares and colleagues maintain that the first step in phenomenological exploration is:

...suspending beliefs or theories about experience, the interviewer raises open questions about the interviewee’s experiences. These

questions lack any categorization or information that could bias the experience of the person. Through the method of open questions, the interviewer aims for the individual to focus their attention on their own experience, reducing the number of possible interpretations (*epoché*). (Olivares *et al.*, 2015, p. 1)

Similarly, Petitmengin and Bitbol (2009, p. 385) emphasize that the interviewer should ask questions ‘which guide the interviewee’s attention towards the various moments and dimensions of his experience, which flag them without suggesting any content’ and they give examples of content-empty questions that ‘enable the researcher to obtain a precise description without instilling her own presuppositions or creating “false memories”’ (Bitbol and Petitmengin, 2017, p. 734). To do this Hurlburt and Heavey (2006) recommend the interviewer ‘bracket’ or set aside his or her presuppositions. Or as Depraz, Varela and Vermersch (2003, p. 27) state:

The most important technique we ask our interviewers to bear in mind — precisely what makes this type of interview useful in research and in professions — is to ask non-leading questions (you can’t whisper the answers under your breath when you’re asking these kinds of questions!) so that interviewees can gain access to their own experience, that is, an eminently pre-reflective material which is not yet conscious, but can be made conscious.

CLI is another method that also emphasizes the content-empty aspect of phenomenological exploration. It ‘aims to minimize co-construction of *content*’ while at the same time recognizing that ‘accounts are co-constructed through the *process* of selecting and asking questions’ (Tosey, 2015, p. 205, emphasis in the original). Content can come from two sources: ‘interviewee content’ comprises all the words used by the interviewee; and ‘interviewer content’ includes any other words that reference a subject or concept introduced into the conversation by the interviewer. The remainder are interrogative words that determine the form of the question. A leading (i.e. non-clean) example from Table 3 will illustrate the distinctions. The interviewer asks: ‘So when it was really productive, how are you deciding how productive it was?’ The words ‘it was really productive’ and ‘productive it was’ are accurate reflections of content provided by the interviewee; ‘you deciding’ is interviewer-introduced content; and ‘So when ... how are ... how ... ?’ provide the interrogative form of the question. An important point is that ‘clean’ questions are not empty of all content, they often contain *interviewee-introduced* content as a way of pointing to a particular aspect of the interviewee’s

experience. However, they are empty of *interviewer-introduced* content (except for some neutral words directly related to the research context as explained in Section 4.2).

3. Clean Language Interviewing

CLI is a systematic research method that facilitates participants to explore the unique micro-structure and micro-dynamics of their experience, as far as possible, from the perspective of the first person. It has been described as a form of linguistic-experiential phenomenology (Owen, 1996) and has been used in a growing number of research topics: how older workers in the Fire & Rescue Service plan for retirement (Pickerden, 2013); managers' work-life balance (Tosey, Lawley and Meese, 2014); the role of knowledge in greening flood protection (Janssen *et al.*, 2014); coaches' evaluation of coaching (Linder-Pelz and Lawley, 2015); tacit knowledge acquisition by novice teachers (Švec, Nehyba and Svojanovský, 2017); leaders' mental models of leadership and leadership development (Cairns-Lee, 2017); narratives from people diagnosed with dementia (Calderwood, 2017); spiritual leadership and self-development within a Sufi spiritual order (Munsoor, 2018); and midwifery decision making (Sanders *et al.*, 2018).

CLI has its roots in the Clean Language psychotherapeutic approach (Grove and Panzer, 1989) which recognizes the central role embodied metaphor plays in cognition and descriptions of experience (Lakoff, 2014). Grove noted that clean questions asked within a clinical context had the following characteristics:

The first objective is for the therapist to keep the language clean and allow the client's language to manifest itself. The second objective is that the clean language used by the therapist be a facilitatory language; in the sense that it will ease entry into the matrix of experience, and into an altered state that may be helpful for the client to internally access his experience... By using Clean Language we uncover the infrastructure of the client's reality... By asking clean questions we shape the location and the direction of the client's search for the answer. In asking a question we do not impose upon the client any value, construct or pre-supposition about what he should answer... We ask our questions so that the client can understand his perspective internally, in his own matrix. (Grove and Panzer, 1989, pp. 8–10)

When talking about *clean* questions we note that the very notion of *cleanness* is a metaphor that can affect our moral judgment (Tobia, Chapman and Stich, 2013). Grove (in Grove and Panzer, 1989) chose

the term to emphasize his desire to not ‘contaminate’ his client’s experience with his own language and presuppositions. In the context of interviewing, we define the concept of *cleanness* as remaining faithful to interviewees’ lexicons, preserving their perspective and logic, and leaving them as free as is possible to answer from their own constructs. We emphasize that a Clean Language interview is different and not necessarily better than other methods.

The CLI method can be delimited with a certain degree of reduction to three rules (Nehyba and Svojanovský, 2017):

- 1) The interviewer makes exclusive use of the literal verbal and non-verbal expressions used by the respondent during the interview.
- 2) The questions asked are designed, as far as possible, to eliminate content assumptions introduced in the words, concepts, and logic of the interviewer.
- 3) The questions facilitate the subject to elaborate on answers that are relevant to the phenomenon under study.

(We use the terms *interviewee*, *respondent*, and *subject* interchangeably.)

Theories of social constructivism point out that it is impossible not to influence in a communication (Gergen, 1999). Therefore inviting answers that are relevant to the phenomenon being studied while remaining ‘clean’ presents a potential dilemma. Clean questions attempt to straddle this dilemma by restricting the interviewer to influencing the process rather than the content of the interview. They do this by encouraging the subject’s attention to stay in his or her field of experience and to describe the phenomenon under investigation from the first person. Therefore, we categorize this method as a second-person interview that supports the interviewee to provide first-person answers. This is indicated in Figure 1 where we place CLI on the sub-scale of the second person close to the first-person perspective. It is debatable whether these perspectives can be displayed on a continuum, side by side, as they may be qualitatively different, but we use this scheme for simplicity.

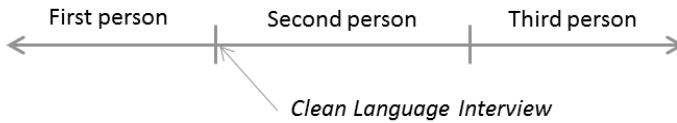


Figure 1. Situating a Clean Language Interview in terms of first-, second-, and third-person perspectives (modified from Depraz, Varela and Vermersch, 2003, p. 85).

3.1. *Literal expressions*

CLI uses the interviewee's precise verbal and non-verbal expressions to facilitate them to 'self-model' their first-person experience (Lawley and Tompkins, 2000). The almost exclusive use of the subject's literal responses may seem mechanical but there is a qualitative difference between viewing the process of Clean Language from the third-person perspective and experiencing it from the first person. If necessary, the apparent strangeness of the questioning can be covered in pre-framing the interview as a method to facilitate introspective searching, as recommended in Descriptive Experience Sampling (Hurlburt, 2011a). However, once the interview is underway, the process usually flows smoothly because the questions are only about the content supplied by the interviewee who is not required to switch attention to constructs and metaphors introduced by the interviewer. We can say that the interviewee naturally concentrates on the phenomenon and is absorbed by it. This fulfils the need to stabilize the interviewee's attention (Petitmengin, 2006). Question syntax — the way in which a researcher composes a question — can help in this respect. In its most formalized form, the syntax of Clean Language consists of three parts: 1) *And* [interviewee's words], 2) *And when/as* [interviewee's words related to an area of their experience], 3) [clean question related to the specified area]. However, depending on the context and the rapport with the respondent, it is not always necessary or appropriate to make use of all three parts of the syntax, and sometimes only the clean question is used.

The repetition and gradual focusing of attention on one aspect of an interviewee's description at a time facilitates him or her to attend closely to their experience which is the subject of the study. Take, for example, the novice teacher who, while being interviewed, says, '*When my teaching starts to go really well I see myself plugged in as all the energy of the pupils flows through me*'. The interviewer starts by acknowledging the interviewee's experience by repeating their

words precisely. The interviewer then invites the interviewee to attend to one part of their description, such as the metaphor ‘plugged in’ and a clean question requests an elaboration of that part of their experience: ‘*And you see yourself plugged in as all the energy of the pupils flows through you. And when you are plugged in, is there anything else about that plugged in?*’ The likely result is that the interviewee attends more closely to the aspect of their experience they have labelled ‘plugged in’ and is able to provide further description (which they may not have been fully aware of before the question was asked). As Lutz and Thompson (2003) note, although subjects are not infallible about their own mental lives they can become aware of important but otherwise tacit aspects of their experience. Furthermore, conscious experience is not only synchronic, it is also diachronic (Petitmengin, 2006). The temporal aspect of experience cannot be elicited through a single question, rather it requires several clean questions which keep the interviewee attending to the sequential nature of a given phenomenon.

The following is a sample of a Clean Language interview (translated from the original Czech) with interviewee-generated words italicized to distinguish them from interviewer-sourced words:

Subject: *When my teaching starts to go really well I see myself plugged in as all the energy of the pupils flows through me.*

Interviewer: *And you see yourself plugged in as all the energy of the pupils flows through you. And when you are plugged in, is there anything else about that plugged in?*

S: *It's like I'm immersed into the current.*

I: *And when you're immersed into the current, what kind of immersed is that?*

S: *I do not think about anything except the pupils and interacting with them, everything follows naturally, and obstacles are opportunities for teaching.*

I: *And you don't think about anything except the pupils and interacting with them, and immersed into the current, and when you're immersed into the current, what happens just before you are immersed?*

S: *I feel I'm opening up to the students. It's like some energy goes from me [touches midline] to them and then back from them to me.*

I: *And you feel you're opening up to the students, and it's like some energy from me [looks at and gestures to interviewee's*

midline] *to them*, and when *energy from me*, where does that *energy* come from?

S: Hmm... [silence] *It starts somewhere inside of me and then it penetrates my whole body until it bursts out.*

I: And *it starts somewhere inside and then it penetrates your whole body until it bursts out*, [pause] and then what happens?

The interviewee's descriptions are replete with metaphors and similes (e.g. '*plugged in*', '*energy of the pupils flows through me*', '*immersed into the current*', '*obstacles*', '*opening up*', '*penetrates*'). They are consistent with Kovecses' analysis (2002) that most metaphors draw on the physiological-material domain to describe more abstract or complex aspects of experience — in other words, a common function of metaphor is reification. In responding, the interviewer accepts the reification and adopts the cognitive linguistic view that metaphor is indispensable to sense-making and description, and that, to a large degree, a person's 'reality itself is defined by metaphor' (Lakoff and Johnson, 2003, p. 272). We do not subscribe to Hurlburt's view that the use of metaphor and simile in cases like this are a form of 'subjunctification' and that 'subjunctifications are signs that expressions are not to be taken at face value' (2011a, p. 117). The Clean Language interviewer assumes that metaphor and simile are used because they 'provide a way of expressing ideas that would be extremely difficult to convey using literal language' (Gibbs, 1994, p. 124). When the interviewee says, '*It's like I'm immersed into the current*', she is saying being '*immersed*' best captures what she wants to describe. The use of metaphor is not an intellectual exercise. *Embodied* metaphors like this not only have psychological meaning, they have physiological and neurological correlates as well (Lakoff, 2014). This is why the question '*...what kind of immersed is that?*' makes sense to the interviewee. By inviting the interviewee to continue to attend to the *particular* experience they call '*immersed*', the interviewer provides the opportunity for the interviewee to deepen the introspection and become aware of that which was previously tacit.

3.2. *What are clean questions?*

Clean questions are designed to eliminate, as much as possible, interviewer-generated content assumptions. These questions avoid both the introduction of new words or metaphors by the interviewer and a leading question structure so as to minimize the potential for contamination of the field of enquiry. It is important to note that *every*

question — including clean questions — assumes something. However, two decades of clinical practice led Grove to develop a set of specific clean questions. These can be applied to a wide range of contexts since they make use of only three general assumptions about the organization of human experience: (1) that each experience has a personal form which enables the subject to distinguish that particular experience from other experiences; (2) that representations of experience are commonly located in a perceptual-mental space (Fauconnier and Lakoff, 2014); and (3) that each event is perceived as part of a sequence of events that happen over time. These methods of organizing experience appear to be common features of all languages (Pinker, 2008). Table 1 lists the ‘classically clean’ questions most frequently used in interviews. There is a clear relationship between the *location* questions and spatial aspects of experience; and the same can be said of the *sequence* questions and temporal arrangements. However, all of the other questions listed are highly flexible and can be equally used to elicit features of the form, spatial, or sequential organization of an interviewee’s inner world. While the classically clean questions suffice for the majority of situations, sometimes a ‘contextually clean’ question is required to be formulated during the interview in order to point attention to a specific aspect of the interviewee’s experience or to the topic of the research.

CLI does not allow the interviewer to paraphrase or interpret the interviewee’s words, and this makes it easier for interviewers to be sensitive to the precise verbal and non-verbal expressions interviewees use to describe their experience. It also helps interviewees to fathom the micro-dynamics and micro-structure of their own experience. Hurlburt (2011a, p. 161) aims to reduce the influence of any specific question by asking multiple deliberately inconsistent questions ‘e.g., “What is your experience at the moment?” “Right then, what were you aware of?” “What if anything presented itself before the footlights of consciousness right then?”’. In contrast, CLI eschews the use of such metaphors, attempting to achieve a similar aim with a single, simple ‘clean’ question.

<p>ATTRIBUTES</p> <ul style="list-style-type: none"> • And what kind of X is that X? • And is there anything else about X?
<p>LOCATION</p> <ul style="list-style-type: none"> • And where/whereabouts is X?
<p>REFLECTIVITY</p> <ul style="list-style-type: none"> • And how do you know X?
<p>METAPHOR</p> <ul style="list-style-type: none"> • And that's X like what?
<p>RELATIONSHIP</p> <ul style="list-style-type: none"> • And when X, what happens to Y? • And is X the same or different as Y?
<p>SEQUENCE - BEFORE</p> <ul style="list-style-type: none"> • And what happens just before X?
<p>SEQUENCE - AFTER</p> <ul style="list-style-type: none"> • And then what happens? • And what happens next?
<p>SOURCE</p> <ul style="list-style-type: none"> • And where does/could X come from?

Table 1. Classically Clean Language Interviewing questions (based on Tosey, Lawley and Mees, 2014). (Note: X and Y are the interviewee's exact words.)

In addition to the precision paid to words, practitioners of Clean Language are encouraged to adjust their paralinguistics: to slow down the speed of delivery, to use soft and curious tonalities with the aim of encouraging introspection, and to restrict their gestures to referencing the other person's iconic or referential gestures in a way that maintains the person's internal perspective (Lawley and Tompkins, 2000). Overall, the CLI approach is another way of 'making sure the interviewer has no knowledge of what the correct details are, or even a knowledge of the range of possibilities' (Froese, Gould and Seth, 2011, p. 58). Minimizing knowledge requires the interviewer to separate his or her inner and outer voice (Vygotsky, 1986) and not

allow that inner voice, which may have a response to the experience of a subject, to affect the content of the questions asked.

The decision about what to attend to in the interview is motivated both by a prior agreement between researcher and interviewee to examine a particular phenomenon, and the interviewee's responses during the interview. It is clear from the construction of clean questions that researchers also seek to let their 'I-ness' fade into the background by not including personal pronouns which reference themselves. The difference between the above examples of clean questions and the traditional interview format can be illustrated by examining a typical request such as: 'I would like you to tell me about your experience.' The flow of attention in this request can be characterized as: 'I would like' (to the interviewer) 'you' (to the subject) 'to tell me' (back to the interviewer) 'about your experience' (back to the subject). CLI questions avoid this shifting of attention and referencing of the interviewer, instead they leave the attention of the subject in his or her perceptual field of experience.

An interesting feature of Clean Language is its ability to reference non-verbal expressions using only the interviewee's non-verbal content, e.g. 'And what kind of [replicate non-verbal expression] is that?' Although little research has made use of this feature, its potential is evidenced by one study that examined the gestures occurring in 96 Clean Language career-coaching sessions (Konat and Juszczuk, 2015). The analysis showed that speakers use specific co-speech gestures to shape complex and abstract concepts such as time, visions of the future, career and life goals. Importantly, the systematic nature of CLI supports interviewers to maintain a consistent method when interviewing several participants and this is even more apparent with projects that involve multiple interviewers.

In summary, clean questions are designed to bracket the researcher's assumptions, to minimize the use of leading questions, and to elicit rich accounts which maximize confidence in the authenticity of the data. The systematic use of the CLI method tends to have a number of effects. It deeply acknowledges and affirms the interviewee's experience and way of describing their subjective reality. The lack of content imposition, interviewer presupposition, distraction, and challenge, along with the preservation of the interviewee's perspective, facilitates him or her to maintain attention on the experience, so they can self-model its dynamic structure and surface tacit knowledge. This can then be reported to the interviewer.

4. Validation of the Second-Person Method

Zumbo and Chan (2014, pp. 9–10) define validation as the process by which ‘we collect and evaluate the evidence to support the appropriateness, meaningfulness, and usefulness of the decisions and inferences that can be made from instrument scores’. Validation in first- and second-person interview methods tends to relate to qualitative rather than quantitative concepts. There is a preference for validation based on the extent to which the respondent reflects, by end-of-interview questions and post-interview conversations that ask subjects to what extent their descriptions correspond to their experience (Bitbol and Petitmengin, 2013b; Hurlburt and Heavey, 2006; Olivares *et al.*, 2015). In these cases, validity is based on credibility in the authority of one’s own experience which the subject is able to assess independently or in collaboration with a researcher.

Nonetheless, the perspective of a third person can also play an important role in evaluating the methods of consciousness research. Systematic third-person measurements can focus on three elements of an interview: the interviewee, the researcher, and the interaction between the two. The ‘experiencing scale’ based on the work of Gendlin (Hendricks, 2009) is an example of a third-person method focusing on the interviewee, and there are now readily available sociometric devices that can be used to record and analyse the interaction between the communicating participants via a range of non-verbal behaviours (Pentland and Heibeck, 2010). In terms of interviewees’ answers, Hurlburt and Schwitzgebel (2007) describe a number of considerations as part of an ‘idiographic validation’. Bitbol and Petitmengin (2013a, p. 29) suggest evaluating performative coherence at ‘several levels of practice: internal coherence in self-assessment and report; interpersonal coherence in dialogue; and triangulated coherence in a network connecting introspective reports with experimental (neurological) practice’. Froese, Gould and Seth (2011) offer an objective ‘double blind’ measurement by a third person for micro-phenomenological and expository interviews. Although Hurlburt (2011b) and Bitbol and Petitmengin (2013b) acknowledge the value of objective measurements to evaluate the reliability of descriptions, they disapprove of the specific protocol suggested by Froese and his colleagues. Hurlburt gets to the heart of the problem when he states, ‘at present, the science of experience has not worked out a method to measure the fidelity of an observation’ (Hurlburt, 2009, p. 187).

4.1. Validity of CLI from a third-person perspective

We recognize validity from first- and second-person perspectives, and we also want to contribute to the development of *complementary* methods of evaluation from a third-party perspective. Below we present a new third-person method of validating second-person interviews by evaluating the relationship between the content provided by the interviewee and the questions asked by the interviewer. The method uses a Cleanliness Rating (Lawley, 2017; Tosey, Lawley and Meese, 2014) which assesses the extent to which the interviewer uses clean rather than leading questions. Although methods for assessing the reliability of answers can be used in conjunction with the Cleanliness Rating, it is important to note that the method does not guarantee the reliability of answers since it is a measure of what the interviewer does.

4.2. Cleanliness Rating

The Cleanliness Rating examines what interviewers aiming to use the CLI method actually do during research interviews. It provides a method of analysing interviewer questions and statements to determine the extent to which they are free of interviewer-introduced content and leading presupposition. The rating is based on the premise that the greater the proportion of clean questions (versus leading questions) the less the interviewee's descriptions are likely to be influenced by (unintended) interviewer imposition, and therefore the more the data gathered are trustworthy first-person accounts. The Cleanliness Rating uses four categories (Classically Clean, Contextually Clean, Mildly Leading, Strongly Leading) as defined in Table 2. Experts in the method (but not involved in the interviews) allocate each question or statement made by the interviewer to one of the four categories. This involves assessing whether the source of the words in the question originates with the interviewee (clean) or the interviewer (leading). This is one way (though still not totally sufficient) to assess whether a question is devoid of the interviewer's conceptualizations and interpretations. In addition, the structure and preconceptions inherent in each question also need to be reviewed. Noticing whispering-an-answer questions requires an understanding of the psycholinguistic effects of metaphor, semantic framing, and presupposition. For example, a question with the form 'Do you [action]?' will be classified as 'leading' since there is growing evidence to support the embodied cognition thesis that comprehension of language

makes use of sensory-motor neuronal networks (Scorolli, 2014; Wilson and Foglia, 2017). In other words, to make sense of (and even to negate) the question requires the interviewee to some degree to ‘take on’ the action in their perceptuomotor systems, and this may unconsciously influence his or her future answers. The subjective element of the method is mitigated by having two or more raters independently assess each transcript. Questions allocated to different categories by the raters are discussed until a consensus rating is agreed. The tabulated results are used to arrive at a summary assessment for each interview.

Classically Clean: uses only the interviewee’s words and questions drawn from the Clean Language question set in Table 1.
Contextually Clean: introduces only <i>neutral</i> words based on the context of the research or the logic inherent in the interviewee’s information.
Mildly Leading: introduces words with the potential to lead but with no discernible effect on the interviewee’s answers.
Strongly Leading: introduces words, especially metaphors, presuppositions, frames, or opinions, that cast doubt on the authorship of interviewee answers.

Table 2. Cleanness Rating definitions (after Lawley, 2017).

The transcript in Table 3 illustrates how the Cleanness Rating can be applied. The interviewee (A) is a female coachee who is evaluating a recent coaching session. The transcript starts after the initial introductions and an agreement about which coaching session will be the subject of the interview.

	Transcript	Cleanness Rating Category and Comment
I:	So how did that session go?	Contextually Clean: Conversational opening question with minimal presupposition.

A:	<i>It went really well... it was a real journey and there was a lot of reflection on the learnings that had been done and that I saw in to some of the patterns of thinking that had been holding me back really clearly and I guess I just felt like I dug deeper into my understanding of the topic we were coaching to and of my own reaction to that, so in that way it was positive for me.</i>	
I:	Right okay, so it was a journey and you saw into your patterns that had been holding you back and you delved deeper and the reflections on learning was the other one. And is there anything else about the session?	Classically Clean: Accurately repeats the interviewee's words and asks for more description using a Clean Language question from Table 1.
A:	<i>Let me think, how it went? I guess it had a, one of the things that I enjoyed about it was that we got into a good rhythm and a good flow in that I often find when I first get in there, kind of like now, sort of not sure what I'm doing and a little bit disorientated just because of the nature of the experience for me, it is disorientating um but then we sort of found a point of focus and used that to move forward to create some specific goals for the session.</i>	
I:	So by the end of the session did you feel that you'd met the goals?	Strongly Leading: The question presupposes the interviewee (a) 'met' her goals, (b) that her way of knowing was to 'feel', and (c) she knew 'by the end of the session' — none of which may have applied. (A cleaner question would have been: And you moved forward to create some specific goals, and then what happened?)

A:	<i>Yes. It was really productive and valuable.</i>	
I:	So when it was really productive, how are you deciding how productive it was?	Mildly Leading: The question leads by pre-supposing the interviewee is ‘deciding’ how productive it was. We can assume she is going through some process but we do not know if the interviewee calls it ‘deciding’. However, the interviewee answers with her own metaphors for her internal process: ‘tools that I used to assess’ and ‘lights go off’. (A cleaner question would have been: And how do you know it was really productive?)
A:	<i>Um, I guess the tools that I used to assess that is: did the lights go off...</i>	

Table 3. Example application of the Cleanness Rating.

5. Method

We examined five research projects which had used the Cleanness Rating as a validation method. Three of the projects were published in academic journals and two were doctoral research. The research topics included: managers’ work–life balance (Tosey, Lawley and Meese, 2014); coaches’ evaluation of coaching (Linder-Pelz and Lawley, 2015); tacit knowledge acquisition by novice teachers (Švec, Nehyba and Svojanovský, 2017); leaders’ mental models of leadership and leadership development (Cairns-Lee, 2017); and curriculum design (Walker, forthcoming). The data consisted of the ratings for 19 interviews conducted by six CLI-trained interviewers and is shown in Table 4.

We noted a variation in how raters counted interviewer interventions that *only* repeated back the interviewer’s words without an overt question. In Project A they were ignored, in Projects C and E they were included under Classically Clean while in Projects B and D they were allocated to their own category. For consistency, all these interventions have been included in the Classically Clean category.

Cleanness Ratings from 19 interviews													
Project:	A		B		C		D		E		Total		Av per interview
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	
No. of interviews	6		6		2		3		2		19		
No. of interviewers	1		1		2		1		1		6		
Classically Clean	214	88	79	28	122	82	239	74	21	50	675	65	36
Contextually Clean	26	11	149	52	20	13	47	14	15	36	257	25	14
Mildly Leading	1	0	46	16	6	4	34	10	6	13	93	9	5
Strongly leading	1	0	12	4	1	1	5	2	1	1	20	2	1
TOTALS	242	100	286	100	149	100	325	100	43	100	1,045	100	55
Clean		99		80		95		88		86		89	
Leading		1		20		5		12		14		11	

Table 4. Cleanness Ratings from 19 interviews (may not sum to totals because of rounding).

Of the 19 interviews, we had access to the individual rater's evaluations in 10 interviews rated by two or three raters. In total 12 raters evaluated 630 questions. The degree to which different raters assign consistent ratings can be used as a measure of the reliability of the data and to indicate the clarity of an instrument's distinctions (Armstrong *et al.*, 1997). To determine the inter-rater reliability we used a number of statistics: a basic measure of Rater Agreement; a Weighted Cohen's kappa for two raters (Landis and Koch, 1977); a Weighted Conger's kappa for more than two raters (Gwet, 2014); and an Intraclass Correlation coefficient (Koo and Li, 2016). The results were generated using the freely available software R (R Core team, 2013) and package 'rel' (Martire, 2017).

6. Results

In the 19 interviews a total of 1,045 interviewer questions and statements were assessed. 65% of these were determined to be Classically Clean, 25% Contextually Clean, 9% Mildly Leading, and 2% Strongly

Leading (rounding to the nearest whole number). The clean/leading ratio showed some variation between projects, ranging from 80/20 to 99/1 with a mean of 89/11. The proportion of Mildly Leading interventions ranged from less than 1% to 16%, averaging five per interview. Most importantly, the mean number of Strongly Leading examples was one per interview.

The inter-rater reliability results of the basic Rater Agreement and the Weighted Cohen's/Conger's kappa calculation are shown in Table 5. The percentage of rater agreement ranged from 50 to 88% for the former test and between 0.50 (a 'moderate' level of reliability using the Landis and Koch kappa Benchmark Scale, 1997) and 0.90 ('almost perfect') in the latter tests.

Frequencies and Reliability Statistics for Interviews					
Interviews	Interviewer / Interviewee*	Number of questions	Marks of Raters	Rater agreement (%)	Weighted Cohen's / Conger's kappa
1	Carl / Mary	42	A,B,C	74	0.74
2	Carl / Glen	60	D,E,F	50	0.51
3	Carl / Julie	58	G,H	62	0.54
4	Carl / Fred	55	I,J,G	75	0.71
5	Carl / Kelly	38	E,G	68	0.62
6	Carl / Grace	53	I,J,G	74	0.75
7	Peter / Charles	69	K,L	83	0.62
8	John / Emma	51	K,L	88	0.54
9	Blanca / Emma	118	K,L	84	0.64
10	Kate / Francis	86	K,L	88	0.90
Total and average	5 / 9	630	12	76	0.66

* Pseudonyms

Table 5. Frequencies and reliability statistics for interviews.

The Cohen's kappa can be misrepresented (Di Eugenio and Glass, 2004). For this reason, the Intraclass Correlation Coefficient (ICC) was calculated. Table 6 shows an average ICC value of 0.72: ranging from 0.48 for one trio of raters ('poor reliability' on Koo and Li's

scale, 2016), to between 0.62 ('moderate') and 0.94 ('excellent') for all the others.

Advanced Reliability Statistics for Interviews			
Interviews	ICC* — 95% CI		
	Estimate	Lower CI	Upper CI
Interview 1 ABC	0.87	0.79	0.92
Interview 2 DEF	0.48	0.32	0.63
Interview 3 GH	0.64	0.46	0.77
Interview 4 IJG	0.83	0.75	0.89
Interview 5 EG	0.68	0.46	0.82
Interview 6 IJG	0.84	0.76	0.90
Interview 7 KL	0.67	0.51	0.78
Interview 8 KL	0.63	0.44	0.77
Interview 9 KL	0.62	0.49	0.72
Interview 10 KL	0.94	0.91	0.96
Average	0.72	0.59	0.82

* Type of ICC: Model — Two-Way Mixed-Effects; Type — Mean of two raters; Definition — Absolute agreement.

Table 6. Advanced reliability statistics for interviews (Intraclass Correlation Coefficient).

Overall, in our view the statistical analysis demonstrates substantial agreement among raters and is comparable with the results obtained for DES by Hurlburt and Heavey (2002).

7. Discussion

Given the claim that second-person interviews can produce authentic first-person accounts, it is incumbent on researchers to 'demonstrate the quality of their work in ways that are commensurate with their assumptions about their use of interviews' (Roulston, 2010, p. 199). Interview fidelity instruments have been developed in the health field (Aggarwal *et al.*, 2014). However, as far as we are aware, the Clean-ness Rating is the first instrument of its kind to attempt to systematically validate *adherence to method* of phenomenological-

type interviewers. For this reason, the results presented here can be used as a benchmark for future researchers.

Our analysis of the 19 interviews shows that, while it is rare for an interviewer to remain 100% clean throughout an entire interview, it is quite possible for well-trained interviewers to adhere to CLI principles 90% of the time. Of the remaining 10%, interviewers can aim to keep their strongly leading questions to no more than one in fifty interventions. Responses to leading questions can be removed from further analysis if the interviewee data are deemed to 'misconstrue their experience' (Froese, Gould and Seth, 2011, p. 47).

Nehyba and Svojanovský (2017) compared the Cleanness Rating of four interviewers: two had gone through an intensive three-day training course and two had attended only a four-hour workshop in the CLI method. Their results showed that the better trained interviewers achieved ratings of 92% and 96% in the *combined* Classically and Contextually Clean categories, while the two less trained interviewers scored 34% and 64%. The difference in competency between well-trained and less well-trained interviewers is consistent with Fowler and Mangione (1990) who found that at least five days of training was required for trainee interviewers to obtain mastery over the questions they asked. It appears that, as well as an evaluation instrument, receiving feedback by rating the cleanness of interviews is a way for researchers to become systematically sensitized (*Einklammerung*) to their own assumptions, models, and metaphors.

The findings of Nehyba and Svojanovský (2017) may be small scale but they match the authors' experience that it is much harder than it appears to consistently ask content-empty questions. A systematic analysis has yet to be conducted, but the numerous examples of leading questions in *model* transcripts provided in academic textbooks and methodological articles suggest that even experienced interviewers are unaware of how often their questions introduce concepts, metaphors, and presuppositions that do not originate with the interviewee. Further research is required to test the authors' hypothesis that interviewers trained in CLI can routinely achieve significantly higher content-empty ratings than is achieved with other interview methodologies.

7.1. Limitations

Clean Language interviews can be effective with a remarkably small number of questions. For example, an analysis of 30 interviews of

European leaders by Cairns-Lee (2017, p. 291) revealed that just four clean questions comprised 69% of all questions asked. This raises the issue of whether the precisely defined questions of CLI unduly limit the phenomenological interviewer. On the one hand, the researcher needs to be free to use a variety of questions in order to address the variability of experience, but on the other hand Stravinsky says, 'My freedom will be so much the greater and more meaningful the more narrowly I limit my field of action' (Stravinsky and Dahl, 1970, p. 65). A comparative analysis could establish whether CLI has different characteristics from other interview methods. However, until there is an accepted method of measuring the quality of data obtained in an interview, it will not be possible to determine whether the restrictions placed on a Clean Language interviewer produce better results than more traditional interview methods.

Using a number of independent qualitative research projects to analyse the consistency of the Cleanness Rating has its limitations. Pooling research data is reliant on the quality of the original studies and the heterogeneity between studies (Ioannidis and Lau, 1999). The research projects quoted in this paper were conducted in four countries and inevitably there will have been some variability of interpretation of the categories among the raters. However, the five studies were chosen because they had all been peer-reviewed and published in one form or another. Clearly a larger pool of results and a more rigorous meta-analysis would provide a greater confidence in the conclusions.

8. Conclusion

The content-empty perspective forms an important feature of the second-person interview. By enhancing interviewers' abilities to apply content-empty theory in practice, we can expect authentic first-person experience to be captured more comprehensively. Clean Language Interviewing is a particularly rigorous method requiring the interviewer to consistently and systematically utilize clean or content-empty questions during a phenomenological interview. CLI adds value to consciousness science through its: specific questions; minimizing the I-ness of the interviewer; unique way of preserving the interviewee's first-person perspective by facilitating them to self-model; and means of maintaining consistency across interviews and interviewers without restricting the exploration of the topic to a fixed format. These features can support the furtherance of second-person psycho-phenomenological research and are in line with Gallagher and

Zahavi's (2008) assertion that the goal of the phenomenological method is to achieve an objective procedure for the research of subjective experience.

Although a number of interview methods have been developed to facilitate first-person descriptions, few have third-person approaches for validating the degree of adherence to the method. This paper introduces a systematic way of evaluating interviews by rating *cleanness* and *leading-ness*. The evaluation of 19 interviews confirmed that Cleanness Ratings of 90% can be consistently achieved by well-trained interviewers who can also restrict *strongly leading* questions to one per interview. These results provide a benchmark for researchers who choose to validate their interviews with the Cleanness Rating. Furthermore, the inter-rater reliability results lend credence to the claim that the Cleanness Rating has taken a step towards 'a method of measuring and calibrating the level of skill of interviewers and interviewees in generating faithful reports of their experience' (Froese, Gould and Seth, 2011, p. 59). The widely varying ability of less well-trained interviewers confirms previous assertions that second-person interviewers require extensive training if they are to ask non-leading or clean questions consistently. The Cleanness Rating presented here offers researchers themselves, and readers of their articles, a degree of confidence that second-person interviews conform to content-empty principles.

References

- Aggarwal, N.K., Glass, A., Tirado, A., Boiler, M., Nicasio, A., Alegría, M., Wall, M. & Lewis-Fernández R. (2014) The development of the DSM-5 Cultural Formulation Interview-Fidelity Instrument (CFI-FI): A pilot study, *Journal of Health Care for the Poor and Underserved*, **25** (3), pp. 1397–1417.
- Armstrong, D., Gosling, A., Weinman, J. & Marteau, T. (1997) The place of inter-rater reliability in qualitative research: An empirical study, *Sociology*, **31** (3), pp. 597–606.
- Bitbol, M. & Petitmengin, C. (2013a) The science of mind as it could have been: About the contingency of the (quasi-)disappearance of introspection in psychology, in Soler, L. & Trizio, E. (eds.) *Science As It Could Have Been*, London: Routledge.
- Bitbol, M. & Petitmengin, C. (2013b) A defense of introspection from within, *Constructivist Foundations*, **8** (3), pp. 269–279.
- Bitbol, M. & Petitmengin, C. (2017) Neurophenomenology and the micro-phenomenological interview, in Velmans, M. & Schneider, S. (eds.) *The Blackwell Companion to Consciousness*, pp. 726–739, Chichester: Wiley & Sons.
- Cairns-Lee, H.M. (2017) *An Exploration of Leadership and its Development through the Inner Worlds of Leaders Using Metaphor*, Doctoral thesis, University of Surrey.

- Calderwood, J.A. (2017) *Pervasive Media and Eudaimonia: Transdisciplinary Research by Practice*, Doctoral thesis, Institute of Creative Technologies, De Montfort University, Leicester.
- Depraz, N., Varela, F.J. & Vermersch, P. (2003) *On Becoming Aware: A Pragmatics of Experiencing*, Amsterdam: John Benjamins Publishing.
- Di Eugenio, B. & Glass, M. (2004) The kappa statistic: A second look, *Computational Linguistics*, **30** (1), pp. 95–101.
- Fauconnier, G. & Lakoff, G. (2014) On metaphor and blending, *Cognitive Semiotics*, **5** (1–2), pp. 393–399.
- Fowler, F.J., Jr. & Mangione, T.W. (1990) *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*, Beverly Hills, CA: Sage Publications.
- Froese, T., Gould, C. & Seth, A.K. (2011) Validating and calibrating first- and second-person methods in the science of consciousness, *Journal of Consciousness Studies*, **18** (2), pp. 38–64.
- Gallagher, S. & Zahavi, D. (2008) *The Phenomenological Mind: An Introduction to the Philosophy of Mind and Cognitive Science*, New York: Routledge.
- Gendlin, E.T. (2004) Introduction to ‘Thinking at the edge’, *The Folio: A Journal for Focusing and Experiential Therapy*, **19**, pp. 1–8.
- Gergen, K.J. (1999) *An Invitation to Social Construction*, London: Sage Publications.
- Gibbs, R.W. (1994) *The Poetics of Mind: Figurative Thought, Language, and Understanding*, Cambridge: Cambridge University Press.
- Grove, D.J. & Panzer, B.I. (1989) *Resolving Traumatic Memories: Metaphors and Symbols in Psychotherapy*, New York: Irvington.
- Gwet, K.L. (2014) *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Multiple Raters*, Gaithersburg: Advanced Analytics.
- Hendricks, M. (2009) Experiencing level: An instance of developing a variable from a first person process so it can be reliably measured and taught, *Journal of Consciousness Studies*, **16** (10–12), pp. 129–155.
- Hurlburt, R.T. (2009) Iteratively apprehending pristine experience, *Journal of Consciousness Studies*, **16** (10–12), pp. 156–188.
- Hurlburt, R.T. (2011a) *Investigating Pristine Inner Experience: Moments of Truth*, Cambridge: Cambridge University Press.
- Hurlburt, R.T. (2011b) Descriptive Experience Sampling, the explication interview, and pristine experience: In response to Froese, Gould, & Seth, *Journal of Consciousness Studies*, **18** (2), pp. 65–78.
- Hurlburt, R.T. & Heavey, C.L. (2002) Interobserver reliability of descriptive experience sampling, *Cognitive Therapy and Research*, **26** (1), pp. 135–142.
- Hurlburt, R.T. & Heavey, C.L. (2006) *Exploring Inner Experience: The Descriptive Experience Sampling Method*, Amsterdam: John Benjamins.
- Hurlburt, R.T. & Schwitzgebel, E. (2007) *Describing Inner Experience? Proponent Meets Skeptic*, Cambridge, MA: MIT Press.
- Ioannidis, J.P. & Lau J. (1999) Pooling research results: Benefits and limitations of meta-analysis, *The Joint Commission Journal on Quality Improvement*, **25** (9), pp. 462–469.
- Janssen S.K.H., Mol, A.P.J., van Tatenhove, J.P.M. & Otter, H.S. (2014) The role of knowledge in greening flood protection: Lessons from the Dutch case study future Afsluitdijk, *Ocean & Coastal Management*, **95**, pp. 219–232.

- Konat, B. & Juszczyk, K. (2015) Multimodal communication in career coaching sessions: Lexical and gestural corpus study, *Empirical Methods in Language Studies*, **37**, pp. 193–211.
- Koo, T.K. & Li, M.Y. (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research, *Journal of Chiropractic Medicine*, **15** (2), pp. 155–163.
- Kovecses, Z. (2002) *Metaphor: A Practical Introduction*, Oxford: Oxford University Press.
- Lakoff, G. (2014) Mapping the brain's metaphor circuitry: Metaphorical thought in everyday reason, *Frontiers in Human Neuroscience*, **8**, art. 958.
- Lakoff, G. & Johnson, M. (2003) *Metaphors We Live By*, Chicago, IL: The University of Chicago Press.
- Landis, J.R. & Koch G.G. (1977) The measurement of observer agreement for categorical data, *Biometrics*, **33** (1), pp. 159–174.
- Lawley, J. (2017) Clean Language Interviewing: Making qualitative research interviews verifiable, in Švec, V., Nehyba, J. & Svojanovský, P. (eds.) *Becoming a Teacher: The Dance between Tacit and Explicit Knowledge*, pp. 115–122, Brno: Masaryk University.
- Lawley, J. & Tompkins, P. (2000) *Metaphors in Mind: Transformation through Symbolic Modelling*, London: The Developing Company Press.
- Lawley, J. & Way, M. (2017) *Insights in Space: How to Use Clean Space to Solve Problems, Generate Ideas and Spark Creativity*, Fareham: Clean Publishing.
- Lifshitz, M., Cusumano, E.P. & Raz, A. (2013) Hypnosis as neurophenomenology, *Frontiers in Human Neuroscience*, **7**, art. 469.
- Linder-Pelz, S. & Lawley, J. (2015) Using Clean Language to explore the subjectivity of coaches' experience and outcomes, *International Coaching Psychology Review*, **10** (2), pp. 161–174.
- Lutz, A. & Thompson, E. (2003) Neurophenomenology integrating subjective experience and brain dynamics in the neuroscience of consciousness, *Journal of Consciousness Studies*, **10** (9–10), pp. 31–52.
- Martire, R.L. (2017) Reliability coefficients, *R package version 0.84.1*, [Online], <https://cran.r-project.org/web/packages/rel/rel.pdf> [10 April 2019].
- Munsoor, M.S. (2018) *Knowing Thyself: A Causal Model of Spiritual Leadership and Self-Development: A Case Study of the Naqshbandiyah Khalidiyah Spiritual Order in Malaysia*, Doctoral thesis, Department of Aqida and Islamic Thought, Academy of Islamic Studies, University of Malaya, Malaysia.
- Nehyba, J. & Svojanovský, P. (2017) Clean Language as a data collection tool, in Švec, V., Nehyba, J. & Svojanovský, P. (eds.) *Becoming a Teacher: The Dance between Tacit and Explicit Knowledge*, pp. 130–147, Brno: Masaryk University.
- Olivares, F.A., Vargas, E., Fuentes, C., Martanez-Perna, D. & Canales-Johnson, A. (2015) Neurophenomenology revisited: Second-person methods for the study of human consciousness, *Frontiers in Psychology*, **6**, pp. 1–12.
- Owen, I.R. (1996) Clean Language: A linguistic-experiential phenomenology, in Tymieniecka, A.T. (ed.) *Analecta Husserliana*, **48**, pp. 271–297, Dordrecht: Kluwer Academic.
- Pentland, A. & Heibeck, T. (2010) *Honest Signals: How They Shape Our World*, Cambridge, MA: MIT Press.
- Petitmengin, C. (2006) Describing one's subjective experience in the second person: An interview method for the science of consciousness, *Phenomenology and the Cognitive Sciences*, **5** (3–4), pp. 229–269.

- Petitmengin, C. & Bitbol, M. (2009) The validity of first-person descriptions as authenticity and coherence, *Journal of Consciousness Studies*, **16** (10–12), pp. 363–404.
- Petitmengin, C. & Bitbol, M. (2011) Let's trust the (skilled) subject! A reply to Froese, Gould and Seth, *Journal of Consciousness Studies*, **18** (2), pp. 90–97.
- Pickerden, A.M. (2013) *How Do Older Workers in the Fire & Rescue Service Deal with Work Life Balance Issues as They Plan For, Approach and Transition Through Retirement?*, Doctoral thesis, Centre for Labour Market Studies School of Management, University of Leicester.
- Pinker, S (2008) *The Stuff of Thought: Language as a Window into Human Nature*, New York: Penguin Books.
- R Core Team (2013) *R: A Language and Environment for Statistical Computing*, Vienna: R Foundation for Statistical Computing, [Online], <http://www.R-project.org/> [10 April 2019].
- Roulston, K. (2010) Considering quality in qualitative interviewing, *Qualitative Research*, **10** (2), pp. 199–228.
- Sanders, J., de Vries, R., Besseling, S. & Nieuwenhuijze, M. (2018) 'Such a waste' — Conflicting communicative roles of Dutch midwifery students in childbirth decision making, *Midwifery*, **64**, pp. 115–121.
- Scorolli, C. (2014) Embodiment and language, in Shapiro, L. (ed.) *The Routledge Handbook of Embodied Cognition*, pp. 127–138, New York: Routledge/Taylor & Francis Group.
- Stravinsky, I. & Dahl, I. (1970) *Poetics of Music: In the Form of Six Lessons*, Cambridge, MA: Harvard University Press.
- Švec, V., Nehyba, J. & Svojanovský, P. (eds.) (2017) *Becoming a Teacher: The Dance between Tacit and Explicit Knowledge*, Brno: Masaryk University.
- Tobia, K., Chapman, G. & Stich, S. (2013) Cleanliness is next to morality, even for philosophers, *Journal of Consciousness Studies*, **20** (11–12), pp. 195–204.
- Tosey, P. (2015) 'And what kind of question is that?' Thinking about the function of questions in qualitative interviewing, in Saunders, M.N.K. & Tosey, P. (eds.) *Handbook of Research Methods on Human Resource Development*, pp. 200–216, Cheltenham: Edward Elgar.
- Tosey, P. & Mathison, J.M. (2009) *Neuro-Linguistic Programming: A Critical Appreciation for Managers and Developers*, Basingstoke: Palgrave Macmillan.
- Tosey, P., Lawley, J. & Meese, R. (2014) Eliciting metaphor through Clean Language: An innovation in qualitative research, *British Journal of Management*, **25** (3), pp. 629–646.
- Varela, F.J. (1996) Neurophenomenology: A methodological remedy for the hard problem, *Journal of Consciousness Studies*, **3** (4), pp. 330–349.
- Vermersch, P. (1999) Introspection as practice, *Journal of Consciousness Studies*, **6** (2–3), pp. 17–42.
- Vygotsky, L. (1986) *Thought and Language*, Cambridge, MA: MIT Press.
- Walker, C. (forthcoming) *Designing Curricula is Like What? Using Clean Language Interviewing to Uncover Metaphors and Models for Designing Curricula*, Doctoral thesis, Liverpool John Moores University.
- Wilson, R.A. & Foglia, L. (2017) Embodied cognition, in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy: Spring 2017 Edition*, [Online], <https://plato.stanford.edu/archives/spr2017/entries/embodied-cognition/> [5 October 2019].

Zumbo, B.D. & Chan, E.K. (2014) *Validity and Validation in Social, Behavioral, and Health Sciences*, vol. 54, New York: Springer International Publishing.

Paper received December 2018; revised October 2019.