



UNIVERSITY OF LEEDS

This is a repository copy of *How to quantify bilingual experience? Findings from a Delphi consensus survey*.

White Rose Research Online URL for this paper:  
<https://eprints.whiterose.ac.uk/187239/>

Version: Accepted Version

---

**Article:**

De Cat, C, Kascelan, D [orcid.org/0000-0002-4465-0844](https://orcid.org/0000-0002-4465-0844), Prevost, P et al. (4 more authors) (Accepted: 2022) How to quantify bilingual experience? Findings from a Delphi consensus survey. *Bilingualism: Language and Cognition*. ISSN 1366-7289 (In Press)

---

This item is protected by copyright. This is an author produced version of an article, accepted for publication in *Bilingualism: Language and Cognition* (<https://www.cambridge.org/core/journals/bilingualism-language-and-cognition>). Uploaded in accordance with the publisher's self-archiving policy.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

Title: How to quantify bilingual experience? Findings from a Delphi consensus survey

Authors' names and affiliations:

1. Cécile De Cat, University of Leeds and Arctic University of Norway
2. Draško Kaščelan, University of Leeds
3. Philippe Prévost, Université de Tours
4. Ludovica Serratrice, University of Reading and Arctic University of Norway
5. Laurie Tuller, Université de Tours
6. Sharon Unsworth, Radboud University
7. The Q-BEx Consortium\*

\*The Q-BEx Consortium: Nouhad Abou Melhem, Ashley Adams, Eva Aguilar-Mediavilla, Nadia Ahufinger, Shanley Allen, Llorenç Andreu, Effrosyni Froso Argyri, Sharon Armon-Lotem, Jacqueline Arsénault, Heather Baer, Colin Baxter, Lisa Bedore, Elma Blom, Mirjam Blumenthal, Ute Bohnacker, Claudine Bowyer-Crane, Krista Byers-Heinlein, Allegra Cattani, Shrivasthi Chakravarty, Vicky Chondrogianni, Dayna Camilleri Clarke, Joanne Cook, Daniel Cubilla-Bonnetier, Ineta Dabasinskiene, Kankan Das, Rachael Davis, Angela de Bruin, Annick De Houwer, Martine Diab, Maaïke Diender, Helen Drummond, Heli Elovaara, Victoria Farrell, Karina Fascinetto-Zago, Giulia Filippi, Caroline Floccia, Max Freeman, Silke Fricke, Margaret Friend, Carolyn Gelenter, Samina Ghafoor, Kleanthes K. Grohmann, Matt Hall, Ewa Haman, Cornelia Hamann, Riikka Härkönen, Jeni Harrison, Carly Hartshorn, Maha Hassan, Erika Hoff, Nayr Ibrahim, Sladjana Indjin, Katy Isaac, Susan Joffe, Holly Joseph, Maria Kambanaros, Kathryn Kashyap, Olga Kepinska, Andrea Kewin, Ekaterini Klepousniotou, Tanja Kupisch, Francesca La Morgia, Gabrielle Lai, Nienke Lam, Jane Le Roux, Mia Le Roux, Annina Manninen, Virginia Marchman, Theodoros Marinis, Prisca Martin, Lourdes Martinez Nieto, Chantal Mayer-Crittenden, Natalia Meir, Liz Metcalfe, Eléonore Morin, Lina Mukhopadhyay, Victoria A. Murphy, Vishnu KK Nair, Marie Newton, Arja Nieminen, Silvia Nieva, Audrey Noel, Ewelina Urszula O'Donnell, Therese O'Sullivan, Ciara O'Toole, Weronika Ozpolat, Felicity Parry, Michelle Pascoe, Rupam Patel, Vrishali Patil Ingle, Sean Pert, Anne-Gaelle Piller, Eva Poort, Christine Potter, Eveliina Rantanen, Saiqa Riasat, Lidia Rodriguez, Wiebke Scharff Rethfeldt, Bethany Faye Schwartz, Irina Sekerina, Miquel Serra Raventós, Saleh Shaalan, Jinder Singh, Beth Skelton, Vicky Slaughter, Anne Margaret Smith, Sini Smolander, Neal Snape, Marina Sokolova, Antonella Sorace, Lisa Spinney-Hutton, Frances Jane Stokes, Luke Swift, Vasim Salim Tamboli, Elena Theodorou, Elin Thordardottir, Ianthi Tsimpli, Rachael Tuckley, Olga Urek, Liza van den Bosch, Klarien van der Linde, Laetitia Vanbruwaene, Josje Verhagen, Virve Vihman, Adriana Weisleder, Gillian Wigglesworth

Note: This version of the manuscript has been accepted for publication in *Bilingualism: Language and Cognition*.

Address of correspondence: Cécile De Cat, University of Leeds, Michael Sadler Building, Woodhouse, Leeds LS2 9JT, United Kingdom; [c.decat@leeds.ac.uk](mailto:c.decat@leeds.ac.uk)

Key words: quantifying bilingualism in children, Delphi consensus survey, researchers, speech and language therapists, teachers

## Abstract

While most investigations of bilingualism document participants' language background, there is an absence of consensus on how to quantify bilingualism. The high number of different language background questionnaires used by researchers and practitioners jeopardises data comparability and cross-pollination between research and practice. Using the Delphi consensus survey method, we asked 132 panellists (researchers, speech and language therapists, teachers) from 29 countries to rate 124 statements on a 5-point agreement scale. Consensus was pre-defined as 75% agreement threshold. After two survey rounds, 79% of statements reached consensus. The need for common measures to quantify bilingualism was acknowledged by 96% of respondents. Agreement was reached to document: language exposure and use, language difficulties, proficiency (when it cannot be assessed directly), education and literacy, input quality, language mixing practices, and attitudes (towards languages and language mixing). We discuss the implications of these findings for the creation of a new tool to quantify bilingual experience.

## 1. Introduction

Most investigations of bilingualism have moved away from classifying participants as bilingual without some documentation of language history and experience. There are several groups of professionals, such as researchers, teachers, and speech and language therapists (SLTs), who often have to document bilingual experience. Both within and across these groups, documenting bilingualism might be required for a range of different purposes: assessing children's development in each language, assessing the risk of a language disorder, assessing their learning potential, assessing their cognitive development, assessing their well-being, etc. This documentation is usually based on language background questionnaires. When studying children, the information is mostly obtained from caregivers, sometimes teachers, or even children themselves.

In a recent review of 48 questionnaires used to quantify bilingualism in children, Kašćelan, Prévost, Serratrice, Tuller, Unsworth and De Cat (2021) observed substantial variation in the documentation of key dimensions of bilingualism, such as language exposure and use, activities in each language, and language skills. For instance, across these questionnaires, exposure and use were documented with particular interlocutors, in specific contexts, during various activities, or as a combination of these. The level of detail varied greatly depending on the interlocutors, contexts, and activities specified in the questionnaire, and depending on whether informants could add categories (e.g., interlocutors) relevant to their circumstances. Language exposure and language use were usually treated as separate constructs but they were not always documented separately, depending on the wording of the relevant questions (which asked, for example, about the frequency of “speaking or hearing” each language). Frequency was usually documented on a Likert scale, but questionnaires used scales of different lengths and with different labels (e.g., quantifying adverbs,

percentage-based points, a combination of frequency adverbs and percentages). Some questionnaires estimated frequencies through open-ended questions. Several questionnaires documented the time that the child spent with each interlocutor, so that the frequency of language exposure from that interlocutor could be adjusted for the proportion of the time spent with the child.

This considerable variability between questionnaires constitutes one of the major hurdles in making the resulting datasets truly comparable.<sup>1</sup> In line with Marian and Hayakawa (2020), we argue in Kašćelan et al. (2021) that bilingualism research is in need of a consensus on what aspects of bilingual experience to document and how. To achieve such a consensus, an inclusive and collaborative approach involving both researchers and practitioners (teachers and SLTs) is required. Such an approach can maximise the potential for cross-pollination between research and practice, and enable a construction of common tools to gather the relevant information about children's bilingualism.

We conducted an international Delphi consensus survey, with the aim of informing the creation of a modular tool for quantifying bilingual experience and achieve consensus between different groups (researchers and practitioners). In doing so, we contribute to the literature by “taking the pulse” of current research and practice, gathering experts' opinions on how bilingualism should be quantified. The survey was completed by 132 researchers, teachers and SLTs from 29 different countries who had worked with bilingual children of various ages and from different bi/multilingual contexts.

The paper is organised as follows. First, we briefly explain the Delphi method and review relevant literature. Second, we outline the methodology adopted in our study, and present the results obtained as well as the list of statements reaching consensus among the

---

<sup>1</sup> Initiatives such as the BLC mini-series (Luk & Esposito, 2020) which aim to gather systematic collections of tools used to document bilingual experiences are a step in the right direction, but they do not aim to enhance the comparability of measures used across bilingualism research and across sectors.

three stakeholder groups. Finally, we discuss the findings and outline future steps towards the creation of a common set of measures to quantify bilingual experience in children.

### ***1.1 The Delphi method***

The Delphi approach is commonly used across disciplines to explore the diversity of opinions among a group of stakeholders or lead them towards a consensus (Iqbal & Papon-Young, 2009). The stakeholders (also known as panellists) are a representative group of experts on the relevant topic of interest. The technique involves an iterative process, in which the panellists express their opinions on a series of statements through two or more survey rounds. The initial set of statements is sometimes itself generated with an open-ended survey (Thangaratinam & Redman, 2005).

The initial set of statements is presented to the panellists via an online survey, in which they are asked to rate their level of agreement with each statement and (optionally) provide comments to justify their ratings. The following round includes a reduced set of statements, depending on the ratings and comments of the previous round. From one round to the next, the statements which have not yet reached agreement or which are in a predefined proximal zone (i.e., grey area) are retained. If required, they can be reformulated. New statements can also be added if necessary.

In the second (and subsequent) round(s), each statement is presented to panellists along with the distribution of responses from the preceding round as well as the panellist's own ratings. This allows them to reflect on their responses in light of group views, and give them the opportunity to maintain or change their rating as they see fit. This process can be

repeated until the consensus is reached, until the predetermined number of rounds is completed, or until it is clear that greater agreement is not possible.

Anonymity is key. It guarantees parity among panellists by giving an equal voice to all, promotes freedom of expression, and limits the risk of bias. The online format allows greater inclusivity, as panellists can be recruited from different geographic areas and complete the survey at their own convenience (within a set time limit).

The design of the present study was informed by guidelines from Hasson, Keeney and McKenna (2000), Iqbal and Pison-Young (2009), and Thangaratinam and Redman (2005), as well as the review by Diamond, Grant, Feldman, Pencharz, Ling Moore and Wales (2014). We modelled parts of our approach on Bishop, Snowling, Thompson, Greenhalgh and the CATALISE consortium (2016), as well as on Langlands, Jorm, Kelly and Kitchener (2008), and Spain and Happé (2019).

## **2. Methodology**

Our aim was to identify the broad consensus on what needs to be documented rather than ask panellists to prioritise which aspects to document (as this would have depended on their particular area of specialism). To maximise the comparability of measures used across studies and practice, we aim to develop a customisable questionnaire allowing professionals to select what is relevant to their purpose from a large set of consensus-informed questionnaire components.

Ethical approval for the study was granted by the University of Leeds. Our procedure was as follows. To generate the first set of statements for the online survey, we organised a workshop with a group of experts. Subsequently, we conducted a two-round online survey in

which a larger group of panellists rated the statements on a 5-point scale, with a possibility to leave comments. It is a common concern in the Delphi approach that repeated iterations can lead to increased panellist attrition (Walker & Selfe, 1996). Therefore, the end of round 2 of the online survey was set as a stopping criterion, no matter the number of statements that reached consensus. This gave panellists the opportunity to change their mind once, and only for those statements in the proximal zone.

In the next two sections, we present the panellist characteristics (section 2.1) and the survey design and procedure (section 2.2). How consensus was defined is explained in the results (section 3.1).

## ***2.1 Panellists/Stakeholders***

The initial workshop (which will be described in the next section) brought together researchers ( $n = 22$ ) and practitioners ( $n = 14$ ) who have worked with bilingual children and extensively used (and often designed) questionnaires to document bilingual experience. We tried to be as inclusive and representative as possible. The workshop participants included experts in typical as well as atypical language development, and had different lengths of experience in bilingualism research. Their expertise spanned different types of bilingual populations, including speakers of heritage languages with various levels of societal prestige, speakers of two majority languages with equal or unequal societal status, and bidialectal speakers. The researchers came from 11 countries: the UK (8), Germany (3), Canada (2), the Netherlands (2), the US (2), France (1), India (1), Israel (1), Poland (1), South Africa (1), and Sweden (1). The practitioners were mostly local (12 from the UK, 1 from Lebanon, and 1 from France), and they ranged from early career to more experienced.



Following the workshop, we used online expression of interest forms (separate ones for researchers, teachers, and speech and language therapists) to expand and diversify the sample of panellists for the online Delphi survey. The expression of interest forms were advertised on our project website and on social media (Facebook, Twitter). We also emailed 247 individual researchers, practitioners, and relevant organisations (e.g., Comité Permanent De Liaison Des Orthophonistes-Logopèdes de l'UE, Audiology and Speech-Language Pathology Associations around the world, the Literacy Association of South Africa, the National Association for Language Development in the Curriculum in the UK).

The expression of interest forms contained questions about the prospective panellists' demographic background and about their experience working with bilingual children. Using these forms, registrations for the survey were submitted by 82 researchers, 61 SLTs, and 27 teachers. We applied the following exclusion criteria: having less than a year of work experience, having never worked with bilinguals, not being able to commit to both rounds of the Delphi survey. This resulted in 13 exclusions (1 researcher, 10 SLTs and 2 teachers), after which 157 stakeholders were retained. In addition, 22 researchers and 14 practitioners from the workshop, as well as 3 researchers who were invited to the workshop but could not attend, were invited to participate in the first round of the Delphi survey. Of the 196 invited panellists, 164 completed round 1 of the survey (response rate: 83%).

Table 1 summarises the distribution of round 1 panellists across self-assigned categories (allowing identification with multiple categories). They came from 30 countries: the UK (52), the US (17), the Netherlands (14), Canada (9), India (7), Spain (7), Finland (6), France (6), Germany (6), South Africa (6), Norway (4), the United Arab Emirates (4), Cyprus (3), Israel (3), Ireland (2), Italy (2), Lithuania (2), Reunion Island (2)<sup>2</sup>,

---

<sup>2</sup> Although Reunion Island is an overseas department and region of the French republic, we counted it separately due to the geographical distance and potentially diverse experiences of the stakeholders in comparison to continental France.

Australia (1), Egypt (1), Estonia (1), Greece (1), Japan (1), Lebanon (1), Malta (1), Mexico (1), Panama (1), Poland (1), Singapore (1), and Sweden (1).

Table 1. *Round 1 panellist breakdown per stakeholder group*

Categories	Sub-categories	Distribution by sub-category, Number (%)	Distribution by category, Number (%)
Researchers		68 (41%)	68 (41%)
Practitioners	Speech and language therapist/ pathologist/ logopedist	38 (23%)	58 (35%)
	Teacher	20 (12%)	
	Speech and language therapist/ pathologist/ logopedist, Teacher	0	
Researchers/ Practitioners	Researcher, Speech and language therapist/ pathologist/ logopedist	24 (15%)	38 (23%)
	Researcher, Teacher	12 (7%)	
	Researcher, Speech and language therapist/ pathologist/ logopedist, Teacher	2 (1%)	
<b>TOTAL</b>			164

In round 2, we invited the 164 panellists who had completed round 1. Of these, 132 completed round 2 (response rate: 80% of the round 2 panellist set, or 67% of the round 1 panellist set). This compares favourably with other Delphi consensus surveys: in a review of 100 Delphi studies, Diamond et al. (2014) reported that only five studies had  $\geq 100$  panellists in the final round, while the rest of the studies either had fewer participants ( $n = 90$  studies) or the number was not specified ( $n = 5$  studies). Table 2 summarises the distribution of round 2 panellists across self-assigned categories (allowing identification with multiple categories). Note that in this round, 10 panellists did not select identical stakeholder labels to describe

themselves as in round 1. Round 2 panellists came from 29 countries: the UK (43), the US (15), the Netherlands (9), Canada (7), Finland (6), Germany (6), Spain (6), France (5), India (5), Cyprus (3), Israel (3), South Africa (3), Ireland (2), Norway (2), Reunion Island (2), United Arab Emirates (2), Australia (1), Egypt (1), Estonia (1), Italy (1), Japan (1), Lebanon (1), Lithuania (1), Malta (1), Mexico (1), Panama (1), Poland (1), Singapore (1), and Sweden (1).

Table 2. *Round 2 panellist breakdown per stakeholder group*

Categories	Sub-categories	Distribution by sub-category, Number (%)	Distribution by category, Number (%)
Researchers	Researcher	57 (43%)	57 (43%)
Practitioners	Speech and language therapist/ pathologist/ logopedist	26 (20%)	40 (30%)
	Teacher	13 (10%)	
	Speech and language therapist/ pathologist/ logopedist, Teacher	1 (1%)	
Researchers/ Practitioners	Researcher, Speech and language therapist/ pathologist/ logopedist	24 (18%)	35 (27%)
	Researcher, Teacher	9 (7%)	
	Researcher, Speech and language therapist/ pathologist/ logopedist, Teacher	2 (2%)	
<b>TOTAL</b>			132

## 2.2 Survey design and procedure

A Delphi consensus survey requires three types of contributors: panellists, moderators and an independent administrator. The panellists included the workshop participants who had informed the initial generation of statements, as well as those recruited through expression of interest forms. The Q-BEx team members (i.e., the main authors of this paper) acted as panellists but also as moderators: they organised and participated in the initial workshop, designed and administered the online survey, and analysed the data. To guarantee anonymity of participation, an independent administrator handled the correspondence with panellists at round 2 (as this included individualised reports). Anonymisation of the data ensured the moderators could not attribute any rating or comment to a particular panellist.

We outline the survey design and procedure below. The full protocol we followed is schematised in Figure 1. An important aim of the design was bias limitation. This was implemented by inviting a group of experts to inform the generation of the initial statements, and by adopting pre-defined procedures for moderation and analysis (as will be explained in the analytic strategy section below). The choice of what topics to include in the Delphi survey was informed by an in-depth review of existing questionnaires (Kaščelan et al., 2021) and by current research and practice (via the workshop).

**WORKSHOP.** The three-day workshop was organised in Leeds in January 2020. The first two days of the workshop were attended by researchers, and the third one by practitioners. One practitioner and five researchers attended all three days. To inform the generation of statements for the online survey, the workshop was organised around thematic presentations by leading experts, a review of the state-of-the-art in bilingual experience questionnaires, issues raised by the participants, and guidance on the principles of questionnaire creation and validation by an expert in psychometrics (Kate Harvey). Throughout the workshop, participants were invited to contribute their views verbally, in writing (using post-its and interactive virtual whiteboards), and through live polling (using

Mentimeter). The thematic presentations focused on the themes in (1), and led to group discussions.

(1) Thematic presentations

- a. Capturing linguistic diversity (Ianthi Tsimpli)
- b. Using experience data to inform the assessment of risk of atypical development (Sharon Armon-Lotem)
- c. Measuring family socioeconomic status in studies of bilingual development (Erika Hoff)<sup>3</sup>
- d. Language mixing (Elma Blom)
- e. Input quality in relation to input quantity (Johanne Paradis)

Emerging themes were identified and discussed, leading to the generation of statements (individually and in small groups). Participants were then invited to contribute what they considered uncontroversial as well as controversial statements.

With the practitioners, the discussions were informed by a critical review of current practice and reflections on practitioners' needs. This also led to the generation of statements. At the end of the workshop, a combined list of 197 statements had been compiled.<sup>4</sup>

Following the workshop, the moderators excluded duplicates and unclear statements. Similar statements were merged, and some were reformulated for clarity, based on our notes from the workshop. This resulted in 53 statements, some of which consisted of several parts (see example (2)). To assess agreement with each part more precisely, every part of an

---

<sup>3</sup> Erika Hoff was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development Grant HD068421.

<sup>4</sup> We generally avoided negative statements, to prevent the complication of double negatives, which are hard to interpret ("I strongly disagree that this is not the case."). The only exceptions were statements s.5, s.13, s.35, s.45 and s.47.

overarching statement was assessed as a separate statement. Therefore, in what follows, we will refer to each part (i.e., sub-component of the original 53 overarching statements) as a separate statement. In this way, altogether, 112 statements<sup>5</sup> were included in round 1 of the Delphi survey.

**PILOT.** The survey was piloted by a group including the six moderators (i.e., the main authors of this paper), as well as six additional researchers and one SLT (the latter seven did not participate in the online study). The aim of the pilot was to check for any errors, assess the clarity of the statements, and optimise their order of presentation.

**DELPHI SURVEY ROUND 1.** The first round of the online survey was administered in April and May 2020. Panellists were given five weeks to complete it (including a one-week extension). In addition to a personalised link to the survey, each panellist was emailed a Briefing Document (see supplement 1) explaining the aim of the study, and a Glossary of the technical terms appearing in the survey (see supplement 2). The panellists were asked to score the statements on a 5-point scale (1 = strongly disagree, 2 = disagree, 3 = I don't know, 4 = agree, 5 = strongly agree). Two statements were different: they asked panellists to indicate the optimal amount of time needed to complete a short and a long version of a bilingual experience questionnaire. Options for a short version included 5 minutes, 10 minutes, and 15 minutes, while options for the long version included 20 minutes, 30 minutes, 40 minutes, 50 minutes, and 60 minutes. There was a space for optional open-ended comments following each statement.

**DELPHI SURVEY ROUND 2.** This round was administered in June and July 2020 and the panellists were given six weeks to complete it (including a one-week extension). Apart from a link to the survey, each panellist was emailed a personalised report of round 1, containing

---

<sup>5</sup> A complete list of round 1 and round 2 statements, panellists' response distribution and comments are available via the Open Science Framework: <https://osf.io/2pd65/>

the distribution of responses for each statement, as well as their own scores for each statement (see supplement 3). In addition, the panellists were emailed a list of round 1 comments (see supplement 4). Finally, the panellists were sent Clarifications and Instructions for completing round 2 of the survey (see supplement 5).

In round 2, the panellists were asked to re-rate 27 of the round 1 statements, and to rate 10 reformulated statements (each one presented immediately following their original formulation), and four new statements. As explained below, we only reformulated statements at round 2 if there was reason to suspect that the lack of consensus at round 1 was due to lack of clarity (determined by participant comments).

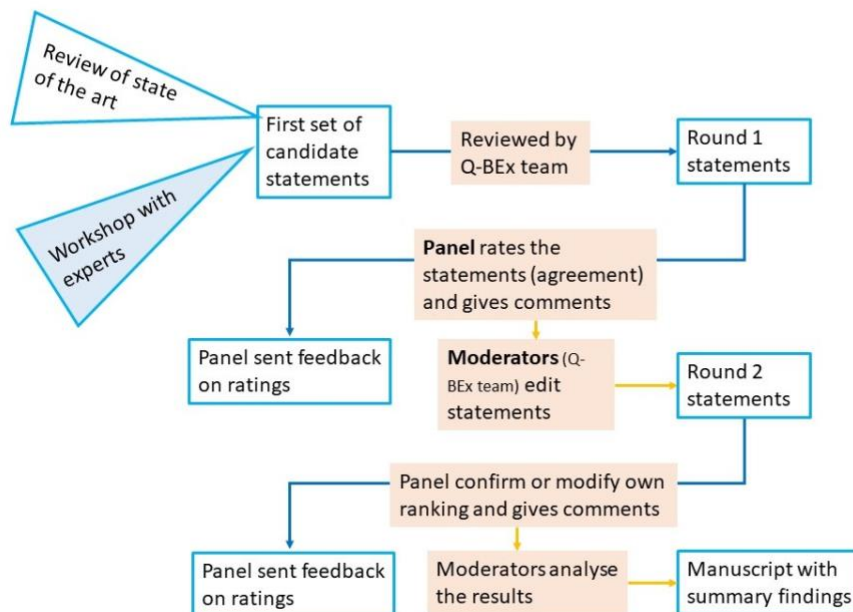


Figure 1. Procedure for the Delphi consensus survey, outlining the role of panellists and moderators<sup>6</sup>

### 3. Results

<sup>6</sup> To illustrate our procedure, we adapted a flowchart Figure 1 from Bishop et al.'s (2016) CATALISE Delphi consensus study.

### ***3.1 Analytic strategy and between-rounds moderation process***

In their guideline paper on the Delphi methodology, Hasson et al. (2000) reported that values between 51% and 80% of agreement have been used as cut-off points for consensus in the literature. In order to determine the consensus cut-off point for our survey, we followed a review by Diamond et al. (2014). This review was based on 100 Delphi studies randomly selected from various disciplines that had been published between 2000 and 2009. Among those using a percentage or a proportion to define consensus, the median consensus threshold was 75%. Consequently, we applied the same criterion in our survey.

(Dis)agreement was defined as follows. Whenever a panellist selected ‘4 = agree’ or ‘5 = strongly agree’, we marked this as agreement with the statement. The round 1 statements reaching an agreement rate  $\geq 75\%$  were considered as having reached consensus and were excluded from round 2 ( $n = 74$  statements), that is, they were immediately included in the final set of agreement-reaching statements (which we analyse in section 3.6). In addition, as an adaptation from Langlands et al. (2008) and Spain and Happé (2019), the round 1 items with an agreement score between 60% and 74%<sup>7</sup> were re-rated in round 2. The purpose of this approach was to reconsider what we refer to as the ‘proximal zone statements’ (i.e., those statements which were not that far from the designated consensus threshold) rather than dismiss them straight away. In round 2, the panellists were invited to reconsider their own scores for these statements, in light of the distribution of round 1 average scores and relevant comments: they could either confirm or modify their score based on this information. Twenty statements met the above ‘proximal zone’ criterion for inclusion in round 2.

A subset of panellists identified themselves as both a researcher and a practitioner ( $n = 38$ ). We refer to these panellists as the ‘dual interest’ group. Since these panellists had

---

<sup>7</sup> The upper limit was 79% in Langlands et al. (2008) and in Spain and Happé (2019).



insights from both perspectives, we decided to give stronger weight to their round 1 ratings in the following way: Statements reaching proximal zone agreement levels (60%-74%) in the dual interest group were automatically included in round 2, even if they were below the 60% threshold in the panel as a whole. Eight statements were thus identified for inclusion to the proximal zone pool (in addition to the 20 mentioned above).<sup>8</sup> Statements reaching an overall agreement rate lower than 60% in round 1 were excluded from round 2 ( $n = 8$ ).

Figure 2 presents the distribution of the statements by agreement rating in round 1. It shows that the pre-defined thresholds for consensus (75% agreement) or proximal zone (60% to 74% agreement) do not correspond to discrete breaks in the agreement distribution. The distribution is skewed (towards high agreement levels) but continuous.

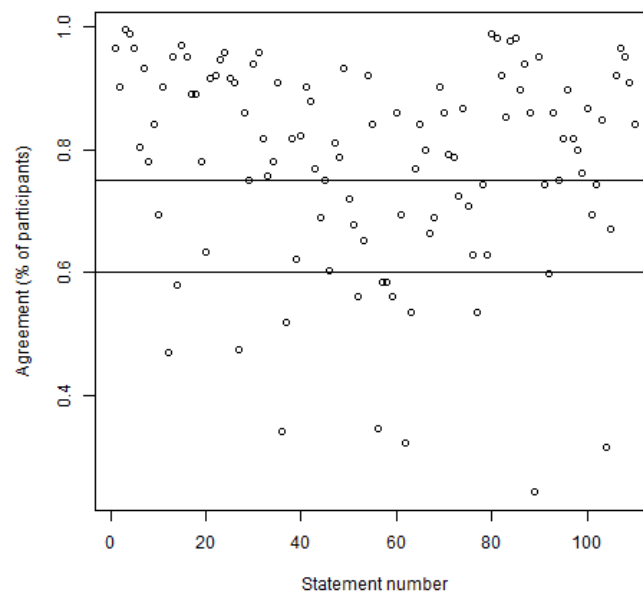


Figure 2. Distribution of round 1 statements by agreement rating (defined as the proportion of participants choosing “agree” or “strongly agree”). The horizontal lines indicate thresholds (for consensus: upper line, and for proximal zone: lower line).

<sup>8</sup> One of these eight statements (statement 39.2) was excluded from round 2 as its related and preceding statement (i.e., statement 39.1) reached a very low overall agreement (32%). Consequently, it did not make sense to include statement 39.2 in round 2 for reconsideration.

Finally, we conducted a thematic analysis of the comments gathered in round 1. This qualitative analysis technique aims to identify the themes that emerge in a dataset inductively. In this approach, a theme is to be understood as a concept central to a mind map of related topics (i.e., a theme is underpinned by a core organising concept, related to other concepts – Clarke & Braun, 2017; Braun & Clarke, 2019). A theme can therefore capture a diversity of inter-connected meanings. The importance of a theme is determined by this web of relationships rather than just frequency of occurrence in the data. The flexibility of this technique allowed us not only to focus on the semantic content of the data, but also to consider the latent levels (i.e., going beyond the semantic content). Consequently, the need for new statements and reformulations could be identified even if they were not explicitly requested. For a further discussion of this approach, see Braun, Clarke and Hayfield (2019).

In round 1, there were 2,486 comments. One of the moderators selected a random subset of 11 statements (1.1, 1.2, 10.1, 10.2, 11, 12, 13, 14, 15, 16.1, 16.2) which in total contained 430 comments (approx. 17% of the data). Based on the reading of these comments, 19 themes were identified and defined inductively. Two other moderators checked the thematic classification of 126 of these comments for consistency, and the list of themes was refined accordingly. The final list of themes and their definitions (provided in supplement 6) were used by all six moderators to classify the remainder of the comments. Each comment could be assigned up to three themes. The comments were then filtered by theme and analysed to identify opinions not yet represented by the survey statements. This resulted in 10 reformulations of the round 1 statements, as well as 4 new statements. Note that all 10 reformulations were rewordings of round 1 proximal zone statements. Consequently, in round 2, the panellists had to re-rate the original round 1 proximal zone statement as well as its reformulation.

In total, round 2 included 27 statements to be re-rated from round 1 (i.e., proximal zone statements based on the overall or the dual interest group ratings), 10 reformulations of the round 1 statements, and four new statements. Round 2 statements reaching an agreement rate  $\geq 75\%$  were considered to have reached a consensus.

### ***3.2 Post-round 2 descriptive analyses***

Overall, there were 55 overarching statements (53 in Round 1 and additional two in Round 2). Many of these consisted of two or more parts (or sub-components), as in (2).

(2) The language(s) used at school should be documented as:

- a. language(s) used by teachers;
- b. language(s) used by the child;
- c. language(s) used by playmates.

For the purposes of our analysis, each part (i.e., sub-component) was counted separately, leading to a total of 126 statements (across the two rounds).<sup>9</sup>

After both rounds, consensus was reached for 79% of statements (98/124). As seen in Figure 3, the distribution of statements varied across agreement bands (with agreement defined as the proportion of panellists expressing agreement or strong agreement for a particular statement). The three highest agreement bands included approximately 85% of statements. By contrast, the 60-70% agreement band included only 6% of statements, and 8% of statements received agreement below 60%. None received less than 20% agreement.

---

<sup>9</sup> Of these, 124 statements were rated on a 5-point agreement scale (strongly disagree, disagree, I don't know, agree, strongly agree). The remaining 2 statements were rated on a time-length scale, as they inquired about preferred time lengths of the short and the long versions of the questionnaire. These 2 statements are excluded from the analyses below. For the distribution of responses on these 2 statements, see supplement 7.

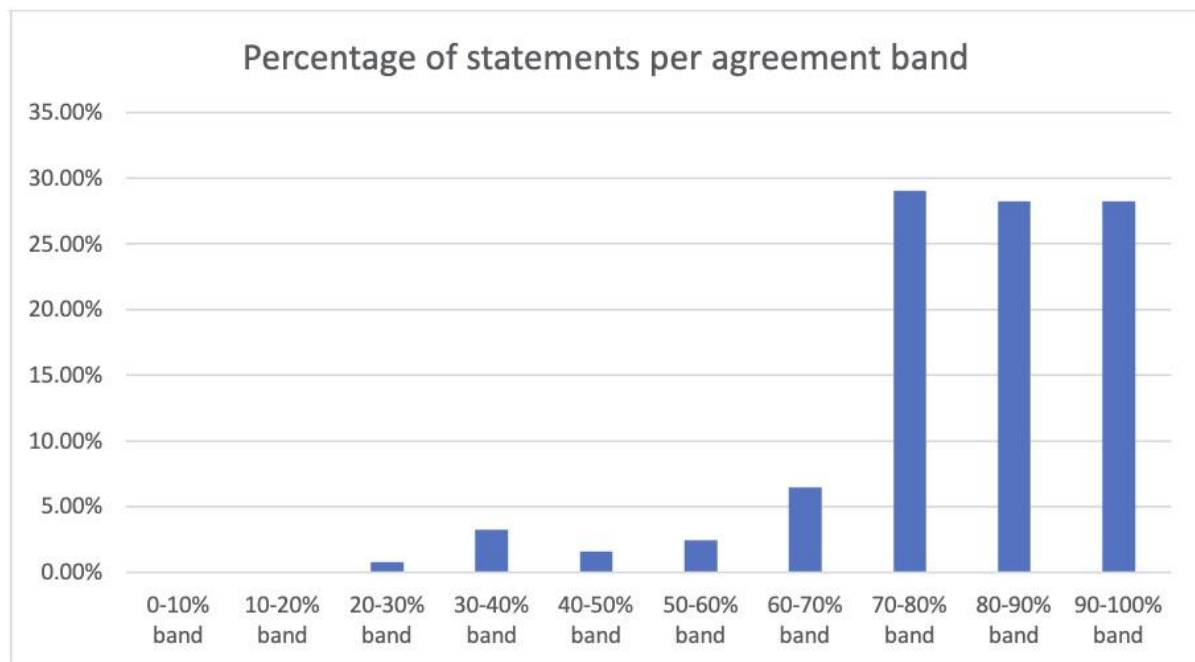


Figure 3. Percentage of statements per agreement band following the two rounds of the Delphi survey

Round 2 included 27 statements in their original formulation, 10 reformulations, and 4 new statements. Most of the original statements (93%, i.e., 25/27) were rated higher in round 2 than in round 1. However, this was not always sufficient for consensus to be reached (requiring a rating as “agree” or “strongly agree” by at least 75% of round 2 panellists). Only 52% of the original statements from round 1 (i.e., 14/27) reached consensus in round 2. From the reformulated statements, 7/10 yielded a higher agreement rating in round 2 than the round 1 originals, while 6/10 reached consensus. Two of these six reformulated statements (statements 31.1 and 49.1) reached consensus both in their round 1 formulation (proximal zone statements) and in their round 2 reformulation. All the new round 2 statements ( $n = 4$ ) reached consensus, bringing the overall round 2 consensus to 59% (24/41).

Figure 4 shows the difference in agreement ratings of proximal zone statements between the rounds. The statements are presented in the ascending order of the difference in

ratings. From the 14 proximal zone statements which reached consensus following round 2, the following three had a jump higher than 15% between the rounds:

- 42. There should be a question on attitudes to language mixing (b) within the local community (including school).
- 5. The questionnaire should not aim to measure the child's language proficiency. This should be done by other means.
- 36.2 [The literacy practices of parents] need to be documented independently of parental education and socioeconomic status.

A visual summary of the entire study can be seen in Figure 5.

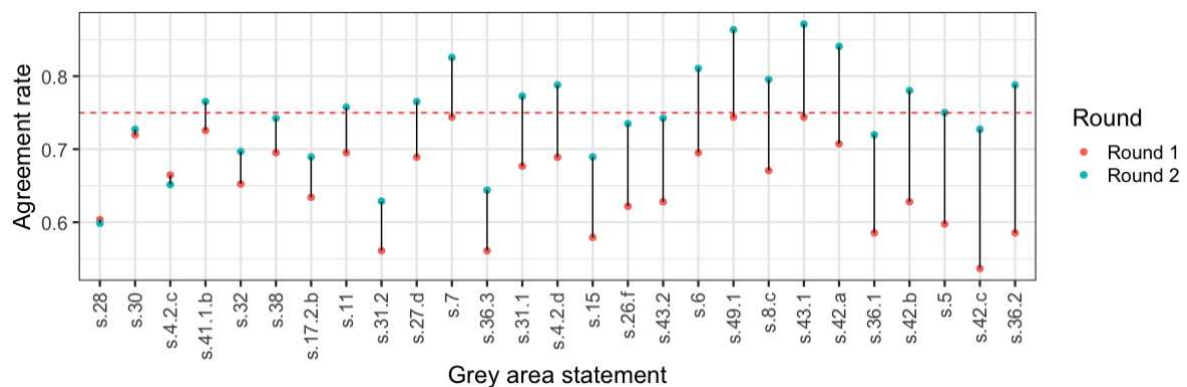


Figure 4. Difference in agreement ratings of proximal zone statements between the rounds (ordered ascendingly by difference size). The dotted horizontal line indicates the consensus threshold.

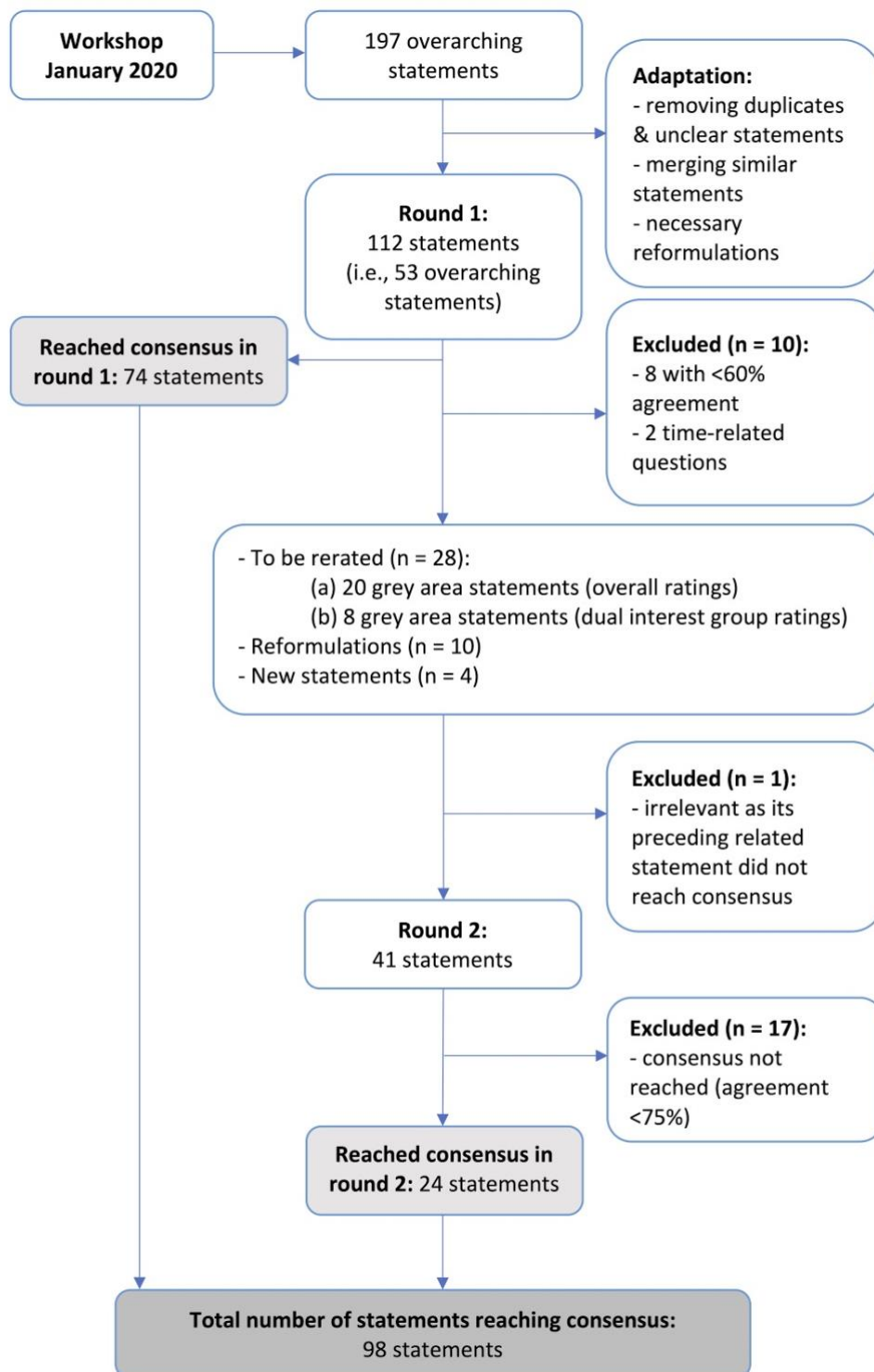


Figure 5. Consensus by stage of the Delphi survey

### 3.3 Attrition analysis

Out of the 164 panellists in round 1, 32 did not respond to round 2. To assess the risk of bias at round 2 as a result of panellist attrition, we reanalysed the data from round 1 by including only responses from panellists who had taken part in both rounds. The results from this subset of panellists differed from the results from the whole round 1 panel in three respects.

First, out of the 74 statements that reached consensus at round 1 in our original analysis, four of them (22, 25, 27e, and 51.1.a) remained marginally below the consensus threshold in the subset analysis (reaching 71.21%, 73.48%, 74.24%, and 72.72% agreement respectively) when the panellists who responded in round 1 only were excluded. While there is no way of knowing whether re-rating these 4 statements in round 2 would pass the 75% consensus threshold, these high levels of agreement (71%-74%) suggest that it is likely that they would. Indeed, as shown in Figure 4, most statements that were re-rated in round 2, increased in the agreement rate by 5% or more.

Second, out of the 28 proximal zone statements identified above, 2 statements (49.1 and 7) already reached consensus at round 1 in the subset analysis (reaching 77.27% and 75% agreement respectively). This is of no consequence, as the re-rating of these statements at round 2 allowed them to reach 86.36% and 82.57% agreement respectively. Finally, statement 31.2 reached the proximal zone in the original analysis, but not in the subset analysis. Again, this is of no consequence as the re-rating of this statement at round 2 did not reach consensus, even in its reformulation (62.87% and 68.18% agreement respectively). Therefore, we conclude that attrition did not increase the risk of bias at round 2.

### ***3.4 Subgroup analysis***

To assess the extent to which the views of researchers aligned with those of the panel as a whole, we analysed the level of consensus among researchers at round 1 (including those from the dual interest group). This revealed that the exact same set of statements reached consensus at round 1 even if only researchers were included. For 92% of statements, there was less than a 5% difference in agreement level between researchers and the panel as a whole. The average difference was 3.4%, and the maximum 10.6%. The direction of the difference was almost evenly balanced (with 53% of statements achieving a higher level of agreement among researchers compared to the panel as a whole, and 47% thus showing a lower level of agreement among the researchers). Differences greater than 5% were always towards greater agreement among researchers than the group as a whole.

### ***3.5 Polarisation analysis***

To ascertain whether there was any polarisation of opinion in the statements that reached consensus, we inspected the distribution of responses for those statements. The data can be seen in Figure S3 (supplement 8), in which statements are ordered according to the level of disagreement (with the would-be controversial statements appearing in the top-most part of the plot). None of the consensus-reaching statement were rated as “strongly disagree” by more than 6% of panellists. We conclude that there was no substantial polarisation with respect to the consensus-reaching statements.

### ***3.6 What is the consensus?***



In this section, we present the 98 statements which reached consensus following both rounds of the Delphi survey (i.e., with an agreement rate  $\geq 75\%$ ). We list them in sections grouped around specific topics, along with the level of agreement reached (expressed as the proportion of panellists who agreed or strongly agreed). We also indicate if the statement was the original round 1 statement, a reformulation, or a new statement added in round 2. For ease of reference, the statement numbers below are as per those in the online survey. A short commentary is provided at the end of some sections as a reflection on the consensus. The complete datasets from both rounds of the online survey can be found via the Open Science Framework: <https://osf.io/2pd65/>

### *Mandate for a new tool*

1.1 There needs to be a set of common measures of children's bilingual language experience, to allow comparability across studies and to facilitate communication across sectors (research, education, therapy). [96%, original]

1.2 These measures should be applicable to children who speak more than two languages. [90%, original]

2. The questionnaire should be accompanied by a tool yielding automatic calculation of objective scores (in each language) of:

(a) current language exposure [95%, original]

(b) current language use [96%, original]

(c) cumulative language exposure [91%, original]

(d) cumulative language use [91%, original]

3. The tool should provide clear guidance about how to interpret the data it produces (where possible). [93%, original]

53. Questions should be as concrete as possible (e.g., asking about daily routines rather than asking the respondent to estimate frequencies in percentages). [87%, original]

### *Language difficulties*

4.1 The questionnaire should contain a section to identify children who might be at risk for a Developmental Language Disorder. [77%, original]

4.2 This should include:

(a) early language milestones [84%, original]

(b) hearing difficulties [80%, original]

(d) issues related to trauma, attachment, prematurity [79%, original]

(e) family history of learning difficulties or speech/language delays [92%, new]

7. The questionnaire should ask about difficulties the child may have (had) with language, in order to identify what might require further assessment by specialists. [83%, original]

55. If the questionnaire includes a section documenting language and/or developmental difficulties, this should under no circumstances be used as a diagnostic tool. [87%, new]

### *Child's proficiency*

5. The questionnaire should not aim to measure the child's language proficiency. This should be done by other means. [75%, original]

6. The child's proficiency should be documented for the language(s) that cannot be tested directly. [81%, original]

The consensus was that questionnaires should document a child's proficiency only when the languages in question cannot be tested directly. Further support for this interpretation comes from the lack of consensus on the reformulation of statement 5 (The questionnaire should not aim to document the child's language proficiency [in any of the languages]. This should be done by other means) and of statement 6 (The child's proficiency should be documented for all their languages).

### *Exposure and use*

#### 8. Exposure and use should be measured (for each language):

- (a) over an average week [85%, original]
- (c) over holiday and school periods separately [80%, original]
- (d) over home and school separately [92%, original]

#### 9.1 The language(s) used at school should be documented as:

- (a) language(s) used by teachers [96%, original]
- (b) language(s) used by the child [95%, original]
- (c) language(s) used by playmates [91%, original]

#### 9.2 Frequencies of use should be documented separately for each type of interlocutor.

[84%, original]

#### 10.1 The languages used outside school should be documented as:

- (a) languages used with parents [99%, original]
- (b) languages used with siblings [99%, original]
- (c) languages used with other carers in the home [96%, original]
- (d) languages used with friends of the family [80%, original]
- (e) languages used amongst the child's friendship groups [93%, original]

- (f) languages used within the neighbourhood [78%, original]
- 10.2 Frequencies of use should be documented separately for each type of interlocutor. [84%, original]
11. The amount of overheard speech (between parents) needs to be estimated. [76%, original]
12. The child's digital language exposure and use needs to be measured (e.g., Internet, social media, gaming). [90%, original]
14. Changes in the child's language exposure over time should be documented. [95%, original]
- 16.1 The child's first exposure to each language should be documented, in terms of:
- (a) age [97%, original]
  - (b) context (e.g., childminder, pre-school, etc.) [95%, original]
- 16.2 A list of potential contexts should be provided. [89%, original]
- 17.1 Time spent in school should be quantified. [89%, original]
- 17.2 This should be done in:
- (a) hours per week [78%, original]

In addition to these statements, it is important to highlight statement 13, which did not reach consensus: The precise measuring of what happens during holidays is unnecessary. It is enough to document whether or not the child travels to the "home" country or has ties with "manifestations of the home country" (e.g., regular contacts online in the home language with dispersed family members). As the statement had a low agreement rate (47%), the documentation of exposure- and use-related practices during holidays might need to be considered.

*Child's education and literacy*

18. Any prolonged period in which the child did not attend formal education should be documented. [91%, original]

19. The questionnaire should ask if the child attended school in another country. [92%, original]

21. The child's frequency of reading in each language should be measured. [86%, original]

22. Child's frequency of writing in each language should be measured. [75%, original]

23. The questionnaire should document any home-language classes that children are attending:

(a) in school [94%, original]

(b) outside school [96%, original]

36.1 Any literacy activities which the parents engage in with the child should be documented. [86%, reformulation]

36.2 This needs to be done independently of parental education and socioeconomic status. [79%, original]

*Input quality*

24. Input quality should be measured as far as this is feasible. [82%, original]

25. There needs to be agreement on a global/composite measure of input quality. [76%, original]

26. The following aspects are indicative of input quality:

(a) parental education [78%, original]

- (b) interlocutor proficiency in each language [91%, original]
- (e) pre-literacy and literacy activities [82%, original]
- (g) playing with peers [82%, original]
- (h) number of interlocutors who interact with the child in each language [86%, new]

27. The language proficiency of the child's interlocutors should be documented (based on the respondent's estimate). This should include estimates for:

- (a) each parent [90%, original]
- (b) any siblings [88%, original]
- (c) other members of the households [77%, original]
- (d) playmates [77%, original]
- (e) teachers [75%, original]

28. The language proficiency of the child's interlocutors should be estimated in relation to typical and representative members of the population/region in which the child lives.

[76%, reformulation]

29.1 The types of activity carried out in each language should be documented (e.g., storytelling, video games, play, etc.). [81%, original]

29.2 A predefined list of activities should be provided to ensure comparability.

[79%, original]

37. Both mother's and father's education need to be documented. [86%, original]

38. If a parent was educated in more than one language, education in each language should be documented separately. [76%, reformulation]

The statements above demonstrate a general requirement across the sectors for estimating child's input quality. While statement 26 outlines the aspects which according to the panellists are indicative of input quality, some of the suggestions did not reach consensus.

These were: interlocutor accent in each language (34%), language mixing (52%), and digital exposure (73%).

### *Language mixing*

31.1 Language mixing should be estimated (in terms of exposure and use). [77%, original]

31.1 Language mixing (heard or produced by the child) should be estimated as part of the child's language exposure and language use. [78%, reformulation]

32. Language mixing that the child is exposed to should be documented separately from the language mixing that the child produces. [77%, reformulation]

33. The questionnaire should ask if the parents use one language in conversation and the child responds in the other. [92%, original]

It is worth noting that in contrast to the statements on language mixing above, where consensus was reached, only about a third of stakeholders agreed with the related statement on mixing in highly multilingual contexts (i.e., In densely multilingual societies, language mixing need not be measured (because it is so prevalent)).

### *Attitudes*

34. There should be a question probing whether the child is unwilling to speak one of his/her languages. [84%, original]

40. There should be a question on attitudes towards each of the child's languages

(a) within the family (at home) [90%, original]

(b) within the local community (including school) [86%, original]

(c) within the broader society [79%, original]

41.1 There should be a question about the status of each of the child's languages

(a) within the local community (including school) [79%, original]

(b) in the "home" country (if applicable) [77%, original]

42. There should be a question on attitudes to language mixing

(a) within the family (at home) [84%, original]

(b) within the local community (including school) [78%, original]

43.1 Parents should be asked if they feel pressurised to speak the societal/majority language.

[87%, original]

### *Background information*

41.2 The child's languages should be identified precisely (e.g., variety, dialect).

[87%, original]

44. The following demographic information should be collected:

(a) child's date of birth [99%, original]

(b) the date of filling in the questionnaire [98%, original]

(c) child's sex [92%, original]

(d) child's birth order [85%, original]

(e) child's arrival into the country (if relevant) [98%, original]

(f) period of the child's life spent in other countries (if relevant) [98%, original]

(g) name of each country in which the child has lived (if relevant) [90%, original]

45. The questionnaire should not label languages as societal, heritage, minority or majority languages. The language names (provided by the respondents) should be used throughout the



questionnaire to identify these languages where relevant. Labels can be applied post-hoc by the researcher or practitioner as required. [94%, original]

46. A range of labels for the child's carers should be allowed by the questionnaire, to better document different family constellations (e.g., other than mother + father).

[86%, original]

In addition to the above, only 24% of the panellists agreed with statement 47 (The questionnaire should not document the immigration history of the child). Such an outcome could be expected as immigration history often complements or is a part of the data on the child's language exposure and use.

### *Questionnaire versions*

48. The questionnaire should be available in an online version, in a paper version, and as an interview protocol. [95%, original]

49.1 The questionnaire should be available in different lengths. [86%, original]

51.1 The questionnaire should be available in different versions for different respondents.

This should include:

(a) a version to be used with 5-7-year-old children [75%, original]

(b) a version to be used with 8-12-year-old children [82%, original]

(c) a version to be used with parents/carers (also usable by adolescents)

[90%, original]

(d) a version to be used with teachers [82%, original]

51.2 The child-focused versions will need to be complemented by a brief parental

questionnaire. [80%, original]

52. The questionnaire in which the child is the respondent should be administered as an interview with the child. [76%, original]

#### *Questionnaire modularity*

49.1 The questionnaire should contain thematic sections (e.g., on language exposure/use, on proficiency, on attitudes, etc.). Each section should be optional, and it should be up to the researchers/practitioners to select which section to use. [87%, reformulation]

50. Some sections of the questionnaire should be optional. The researcher/practitioner should be able to exclude some sections. [86%, original]

54. Some parts of the questionnaire should not be administered if asking these specific questions is not adequate (for safety, political, personal, or any other ethical or relevance reasons). [88%, new]

## **4. Discussion**

This Delphi consensus survey aimed to ascertain the level of agreement among researchers and practitioners regarding how bilingual experience should be documented. Our aim was to identify the broadest consensus possible, in order to inform the design of a new, customisable questionnaire allowing researchers and practitioners to select the components relevant to their purpose.

The round 1 statements were formulated based on the outcomes of an exploratory workshop with 22 researchers and 14 practitioners. In two rounds of the online survey, a diverse set of 132 panellists from 29 countries rated a total of 126 statements<sup>10</sup> (112 original

---

<sup>10</sup> Two of these were not rated on an agreement scale and were excluded from the consensus calculations below.

statements, four new statements, and 10 statement reformulations). Furthermore, 27 original round 1 statements were re-rated in round 2. Several measures were adopted to limit the risk of bias: we used a diversification strategy for the recruitment of panellists; the data was anonymised upon collection; we used pre-defined criteria for consensus and for selecting statements to be re-rated; and we carried out post-hoc bias analyses. While it is impossible to avoid bias entirely, we believe these measures give credence to our approach. We are confident we have captured a range of opinions reflecting a variety of centres of interests among an international community of researchers and practitioners.

Overall, the level of consensus was high: 79% of statements (i.e., 98/124) reached consensus across rounds, and only 8% of statements (10/124) received less than a 60% agreement rate. This suggests that the views of the experts from the initial workshop were fairly representative of those of the wider, more diverse panel participating in the online survey, including a solid proportion of uncontroversial views but also some more controversial ones.

#### ***4.1 The consensus***

There was an almost unanimous call for a set of common measures of bilingual experience, enabling greater comparability across studies, and facilitating exchanges and cross-pollination across sectors. Also, almost unanimous was the desire for an automated calculator of language exposure and language use, and the need to allow for measuring multilingual (not only bilingual) experience.

Consensus was also reached regarding the need to document a number of aspects of bilingual experience. This includes, for each language: language exposure and use, language mixing, language difficulties experienced, proficiency (if it cannot be assessed by other

means), education and literacy, indices of input quality, language mixing practices, and attitudes (towards each language and towards language mixing). Consensus levels were the highest in relation to language exposure and use, and the need to document them in detail (i.e., across interlocutors, in different contexts, and over time).

The variability observed with respect to other aspects is likely a reflection of the fact that they have hitherto been researched less systematically. The survey results indicate that researchers across the board agree these aspects are likely to be important but require more supporting evidence and/or more scrutiny. For example, input quality and language mixing are starting to attract more attention in bilingualism research (see, e.g., Unsworth, 2016 and Byers-Heinlein, 2013, respectively), and this is reflected in a set of quality-related and language mixing statements which reached consensus. Nonetheless, several statements relating to the documentation of language mixing did not pass the consensus threshold. For instance, there was no agreement about the documentation of language mixing for each interlocutor in interactions with the child (statement 30); nor was consensus reached on documenting language mixing through examples of different types of mixing or their frequency (statement 31.2 and its reformulation). This might reflect scepticism as to whether questionnaires can reliably document these aspects of language mixing. Further research will be necessary to elucidate these points.

Another aspect eliciting diverse reactions was language proficiency. Round 2 reformulations revealed that this was due to a caveat: if language proficiency could not be assessed by other means, the agreement was that it should be documented via a questionnaire. Similarly, as long as the questionnaire is not seen as a diagnostic tool, the consensus was that the child's difficulties with language should be documented.

The need for a flexible or modular questionnaire elicited strong consensus. The constant tension between the level of detail aspired to and the constraints inherent to data

collection will often result in having to forgo the documentation of some aspects. A modular questionnaire will allow for this. The exclusion of some aspects will also partly depend on foci of interest and the purpose of documentation (e.g., screening for language difficulties vs. informing a study on a particular aspect of bilingual experience). Different versions of the questionnaire were also considered necessary to adapt to different types of respondents (e.g., caregiver, child, teacher).

The apparent contradiction between the call for a flexible and modular tool and the acknowledgement of the necessity to document a wide range of aspects of bilingual experience brings us to the question of what is ‘core’ vs. what can be considered optional in the documentation. While this question was not asked directly in the survey, we believe variations in level of consensus can be interpreted as useful indicators. Language exposure and use, as well as some indicators of language difficulties and some indicators of input quality, thus seem to emerge as core aspects of the quantification of bilingual experience. Ultimately, though, the identification of an essential ‘core’ is an empirical question, which will require comparing data from large and diverse groups of bilinguals and multilinguals, using identical measures. And this empirical investigation will need to ascertain not only which aspects of bilingual experience are part of the ‘core’, but also the minimum level of detail required to measure (or document) them reliably. Indeed, the cognitive load involved in completing bilingual experience questionnaires can be quite complex, as respondents are required not only to recall language practices over long periods of time and many different contexts, but also to estimate their frequency. In addition, they might not have been direct witnesses of these practices (e.g., in the case of a parent estimating what happens at school), or they might not have been fully aware of them (e.g., in the case of language mixing). This is likely to result in unavoidably high levels of error in the measurements. It is therefore not necessarily the case that a more detailed questionnaire will elicit more precise information.

Here too, empirical investigation will be necessary to identify the optimal level of detail to be targeted by bilingual experience questionnaires.

## ***4.2 Limitations***

While the Delphi method provides a flexible approach to accommodate the needs of various fields, a clear limitation is the lack of agreed standards on what should count as consensus and how it should be interpreted (Iqbal & Papon-Young, 2009). To limit the risk of bias, we adopted a pre-defined criterion informed by Diamond et al.'s (2014) review, which identified 75% agreement among panellists as the median consensus threshold among the publications in which consensus was defined as a percentage or proportion. Furthermore, we adopted a pre-defined criterion to identify round 1 statements in the proximal agreement zone (i.e., the proximal zone statements): informed by the approach of Langlands et al. (2008) and Spain and Happé (2019), any statement reaching agreement between 60% and 74% of round 1 panellists was automatically selected for re-rating at round 2. We also pre-defined a dual interest group (identifying as both researchers and practitioners), whose proximal zone statements (i.e., statements with the agreement rate 60%-74%) were automatically selected for round 2, in addition to those of the panel as a whole. These procedures enabled us to mitigate the risk of a potentially too conservative consensus threshold.

In spite of our diversification strategy, the panel remained predominantly Western-centric (both in the workshop and in the online survey). The use of English and the use of an online platform to conduct the study aimed to increase inclusivity, but at the same time it discriminated against non-English speaking stakeholders and individuals without access to the internet. Future work should therefore seek to improve representation, as many significant voices have likely not been included.

We also note that there are no validated quality indicators of Delphi studies. However, Diamond et al. (2014) proposed a set of four elements to include in Delphi publications to increase their value: (1) provide reproducible criteria for panellist selection, (2) state the number of rounds performed, (3) provide clear criteria for dropping the items, and (4) clarify whether there is a stopping criteria other than the number of rounds specified. We reported on each of them in this manuscript.

The ultimate limitation of the Delphi method is that, in spite of all attempts at bias limitation, it remains possible that the correct answer or opinion was not identified (Hasson et al., 2000; Iqbal & Pipon-Young, 2009). Opinion is also likely to evolve as bilingualism research progresses. However, we believe the findings of this survey could enable a step-change in bilingualism research through the adoption of a common method to document bilingual experience, thereby enabling greater comparability across studies (and across populations), and increased synergy between practitioners and researchers. The validity of the tool developed on the basis of the Delphi consensus will need to be assessed based on new empirical evidence. It may turn out that some of the aspects of bilingual experience which were deemed necessary to document (as per the consensus reached by the Delphi survey) can in fact not be reliably documented via questionnaires. If that is the case, it will indicate that opinions need to change. This paper represents the first step in this long process, that is, the documentation of current opinions.

There are also limitations inherent to the quantification of (aspects of) the bilingual experience, which was at the heart of this Delphi survey. Such an approach might not be sufficient to capture key aspects of bilingualism associated with variation across socio-cultural contexts, such as non-industrialised countries without a robust education system. Ethnographic approaches (not considered in this survey) might offer insightful alternatives in some contexts. It is possible that standardised questionnaires are inadequate to document

bilingualism in some populations. The extent to which bi/multilingualism in these populations can be conceptualised along the same lines as in other populations is an important question beyond the scope of this paper and will require further research.

### ***4.3 Next steps***

The main authors of this paper are currently designing an online questionnaire<sup>11</sup> and background calculator meeting the requirements identified via the Delphi consensus survey and informed by a review of the state of the art in bilingualism questionnaires (Kaščelan et al., 2021), and by methodological insights from the literature on psychometrics (e.g., DeVellis, 2017). This questionnaire will allow some level of customisability, so that professionals can choose the components relevant to their research objectives, clinical practice, or educational aims. To limit the burden on respondents, we will also carry out a cost-benefit analysis aiming to identify the optimal balance between informativity and error margin. This is particularly important as lengthy questionnaires are likely to be too burdensome for some marginalised communities, which would result in less representative population samples. As is the case with the creation of any standardised tool, the domain of applicability of this new questionnaire will need to be assessed empirically. In the long term, a multi-team approach will be necessary for full validation and optimisation.

The creation of a new tool raises several methodological challenges. First, the constructs that make up bilingual experience need to be operationalised precisely and concretely to avoid reification. Second, there is a risk of normalisation inherent to the creation of a standard questionnaire. This risk must be limited by embedding into the questionnaire design

---

<sup>11</sup> The design phase took place while the manuscript for this paper was under review. The new online tool is now available for free at <https://q-bex.org>.



the intention to capture the diversity of bilingual experiences. This will however need to be balanced with the need to keep the questionnaire sufficiently short and clear for respondents.

## **5. Concluding remarks**

The need for standardisation in how bilingualism is characterised and categorised has become uncontroversial. This Delphi consensus survey has highlighted the readiness of bilingualism researchers and practitioners from both clinical and educational settings to adopt common methods for the documentation of bilingual experience, to enhance the generalisability of research findings and facilitate exchanges between research and practice.

Several new profiling measures have been proposed recently, as global indices of bilingualism: for example, the LSBQ composite score (Anderson, Mak, Keyvani Chahi & Bialystok, 2018; Anderson, Hawrylewicz & Bialystok, 2020), language entropy (Gullifer & Titone, 2020), or a possible bilingualism quotient (Marian & Hayakawa, 2020). Independent of their conceptual validity, the reliability of these measures will be determined by how their components are documented and quantified. This in turn will require comparability of measures, which will only be possible if similar documentation tools are used across research teams. The results of this Delphi consensus survey constitute a first step in that direction.

Data availability statement: The data that supports the findings of this study are openly available via the Open Science Framework at <https://osf.io/2pd65/>

Competing interests declaration: The authors declare none.

## References:

Anderson, J. A. E., Hawrylewicz, K., & Bialystok, E. (2020). Who is bilingual? Snapshots across the lifespan. *Bilingualism: Language and Cognition*, 23, 929–937.

<https://doi.org/10.1017/S1366728918000950>

Anderson, J. A. E., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, 50, 250–263. <https://doi.org/10.3758/s13428-017-0867-9>

Byers-Heinlein, K. (2013). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism: Language and Cognition*, 16, 32-48. <https://doi.org/10.1017/S1366728912000120>

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE consortium (2016). CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children. *PLoS ONE*, 11(7), 1-26.

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589-597, DOI:

10.1080/2159676X.2019.1628806

Braun, V., Clarke, V., & Hayfield, N. (2019). 'A starting point for your journey, not a map': Nikki Hayfield in conversation with Virginia Braun and Victoria Clarke about thematic analysis. *Qualitative Research in Psychology*, DOI: 10.1080/14780887.2019.1670765

Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology*, 12(3), 297-298, DOI: 10.1080/17439760.2016.1262613

DeVellis, R. F. (2017). *Scale Development Theory and Applications* (4th ed.). SAGE Publications, Inc.

Diamond, I. R., Grant, R. C., Friedman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, 67, 401-409.

Gullifer, J. W., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, 23(2), 283–294.

<https://doi.org/10.1017/S1366728919000026>

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, 32(4), 1008-1015.

Iqbal, S., & Pilon-Young, L. (2009). The Delphi method. *Methods*, 22(7), 598-601.

Kaščelan, D., Prévost, P., Serratrice, L., Tuller, L., Unsworth, S., & De Cat, C. (2021). A review of questionnaires quantifying bilingual experience in children: Do they document the same constructs? *Bilingualism: Language and Cognition*, 1–13.

<https://doi.org/10.1017/S1366728921000390>

Langlands, R. L., Jorm, A. F., Kelly, C. M., & Kitchener, B. A. (2008). *Journal of Affective Disorders*, 105, 157-165.

Luk, G. & Esposito, A. (2020). BLC mini series: Tools to document bilingual experiences. *Bilingualism: Language and Cognition*, 23(5), 927-928.

Marian, V., & Hayakawa, S. (2020). Measuring bilingualism: the quest for a “bilingualism quotient”. *Applied Psycholinguistics*, 1-22. doi:10.1017/S0142716420000533

Spain, D., & Happé, F. (2019). How to Optimise Cognitive Behaviour Therapy (CBT) for People with Autism Spectrum Disorders (ASD): A Delphi Study. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*. <https://doi.org/10.1007/s10942-019-00335-1>

Thangaratinam, S., & Redman, C. (2005). The Delphi technique. *The Obstetrician & Gynaecologist*, 7(2), 120-125.

Unsworth, S. (2016). Quantity and Quality of Language Input in Bilingual Language Development. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism Across the Lifespan: Factors Moderating Language Proficiency* (pp. 103-121). De Gruyter Mouton.

Walker, A. M., & Selfe, J. (1996). The Delphi technique: a useful tool for the allied health researcher. *British Journal of Therapy and Rehabilitation*, 3, 677–680.

## **Supplement 1**

**Paper title:** How to Quantify Bilingual Experience? Findings from a Delphi Consensus Survey

**Content:** A briefing document sent to the panellists participating in round 1 of the online survey.



16 March 2020

**Delphi Panel Briefing Document**  
Quantifying Bilingual Experience: A Delphi Consensus Study

You are being invited to take part in the research project 'Quantifying Bilingual Experience: A Delphi Consensus Study'. Before you decide to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information (contact details can be found below). You will have two weeks to decide if you want to take part in this study.

**What is the purpose of the project?**

This study aims to establish an agreement among key stakeholders regarding how bilingual language experience should be quantified, taking into account the needs of researchers, teachers, and speech and language therapists (SLTs). The consensus will be sought by running a **Delphi consensus survey**. For a detailed description of the Delphi methodology see <https://thepsychologist.bps.org.uk/volume-22/edition-7/delphi-method>

**Why have I been chosen?**

The Delphi survey method requires a group of panellists/stakeholders to rate statements on a particular topic (in this case quantifying bilingual experience) in several rounds (in this case two) with the aim to achieve agreement wherever possible. As you are a member of one of the stakeholder groups identified above, you have been chosen to participate in this study.

**Do I have to take part?**

It is up to you to decide whether or not to take part. If you do decide to take part, you will be asked to give your informed consent electronically, at the beginning of the online survey form. You can withdraw at any time without it affecting any benefits that you are entitled to in any way. You do not have to give a reason for withdrawal.

**What do I have to do?**

If you decide to participate, you will take part in a two-round online survey.

In the first round of the online survey, you will be asked to rate a series of statements on quantifying bilingual experience. You will have to rate them on a Likert scale based on how much you agree with them. You will also be given a chance to leave a comment on each statement. The online survey will take about 30-60 minutes to complete and the completion deadline is 30 April 2020 (23:59 BST).

In the second round of the online survey, you will be sent a report of the previous round containing the distribution of ratings for each statement (together with anonymised comments), along with your own previous ratings. You will be asked to rate the statements again (either keeping your original responses or changing them in light of other panellists' responses). You will have a chance to justify your choice and leave a comment or respond to any of the comments made previously. This survey will take about 30-60 minutes to complete, and you will be given about 4 weeks to do so.



## UNIVERSITY OF LEEDS

At the end of the study (expected end date: 30 September 2020), you will receive the final report explaining where consensus has been reached and where it has not.

### **What are the possible disadvantages and risks of taking part?**

To the best of our knowledge, there are no known disadvantages or risks of taking part.

### **What are the possible benefits of taking part?**

Participating in this project could benefit your future practice as a stakeholder. In a follow-up project, the outcome of this study will be used to inform creation of a questionnaire to quantify bilingual experience, which will be open access and made available in 13 languages. Considering that it will be based on the consensus reached between an international group of researchers, teachers, and SLTs, this tool will be relevant to all the stakeholders participating in our current study.

### **Use, dissemination and storage of research data**

Your data will be anonymised by removing any personal information that could identify you. An independent researcher will be recruited to process your unanonymised data. This data will be stored on his/her University drive (accessible only by them) until the end of the project. The independent researcher will anonymise your data and forward it to the research team for the analysis. Anonymised data will be indefinitely kept on a protected folder in OneDrive, shared between the research team.

### **What will happen to my personal information?**

At the end of the project, the unanonymised data, and your personal information will be permanently deleted.

### **What will happen to the results of the research project?**

We will write the final report and an open access publication, both of which will contain a full dataset in the anonymised form. This will be made available to anyone indefinitely. The reports and the publication with the anonymised data will be stored in online databases, on the project website, and they will be shared with other stakeholders or interested individuals. The results of the study will be presented at conferences always in the anonymised form. You (as a panellist) will be listed in the publication (as part of the consortium of authors) unless you inform us otherwise. No link will be made between individual ratings/comments and any of the panellists.

Even though you are allowed to withdraw from the study at any point, once the data is anonymised and attached in reports/publication, it will not be possible to delete it.

### **Who is organising/funding the research?**

This study is part of a project funded by an ESRC grant awarded to Prof. Cécile De Cat (University of Leeds).

### **Contact for further information**

If you have any questions or issues that you would like to discuss, please contact us via [qbex@leeds.ac.uk](mailto:qbex@leeds.ac.uk) or by writing directly to the Principal Investigator, Prof. De Cat, at [c.decat@leeds.ac.uk](mailto:c.decat@leeds.ac.uk).

Thank you for taking time to read through the project briefing!

## **Supplement 2**

**Paper title:** How to Quantify Bilingual Experience? Findings from a Delphi Consensus Survey

**Content:** A glossary of technical terms sent to the panellists participating in round 1 of the online survey.



## Quantifying Bilingual Experience (Q-BEx) Delphi Consensus Study Glossary

**Cumulative language exposure** is the sum of bilingual's exposure to each language over time.

**Cumulative language use** is the sum of bilingual's use of each language over time.

**Current language exposure** is the present frequency of exposure to each language.

**Current language use** is the present frequency of using each language.

**Densely multilingual societies** are societies in which languages are frequently mixed and rarely used separately.

**Heritage language** is the language of home or a community that differs from the societal (i.e. majority) language and it is often the language of immigrant or indigenous communities. Sometimes also called "home language".

**Home country** here refers to a bilingual's or their caregiver's country of birth if different from the country of current residence.

**Home language** refers to a language of bilinguals that is not the societal language and is used in their home (although the frequency of its use in the house varies depending on the family). For many bilinguals, a home language is their heritage language. Sometimes also called "heritage language".

**Input** refers to the language(s) that the child is exposed to.

**Language exposure** refers to the language(s) that the child hears and/or reads.

**Language mixing** is the use of more than one language in a conversation or in writing. We use this term broadly to include what linguists refer to as code-switching or code-mixing.

**Language status** refers to the prestige that a language has within a society.

**Language use** refers to the language(s) that the child produces by speaking and/or writing.

**Majority language** is the dominant language of a wider society. It is also referred to as the societal language (see below).

**Minority language** is bilingual's home or heritage language which is not dominant in the society where they live. In some situations, it equates to the heritage language or home language.

**Output** refers to the language(s) that the child uses.

**Overheard speech** refers to the language(s) that the child is exposed to even when not spoken to directly. For instance, this would include scenarios when the child is in the same room with caregivers who talk to each other (and not to the child). In this case, the child is still exposed to language(s) although s/he is not involved in the conversation.

**Societal language** refers to a language used by the wider society in a country. For instance, in England, English is the societal language.

## **Supplement 3**

**Paper title:** How to Quantify Bilingual Experience? Findings from a Delphi Consensus Survey

**Content:** An example of a personalised round 1 report sent to the panellists with an invitation to complete round 2 of the online survey.

# Quantifying Bilingual Experience (Q-BEx) Delphi Consensus Study

## Round 1 Report: Overall Percentage Distribution

Panellist ID: [REDACTED]

**Statement 4.1:** The questionnaire should contain a section to identify children who might be at risk for a Developmental Language Disorder. [Agreement: 77%]

**Statement 4.2:** This should include:

**4.2.a: early language milestones** [Agreement: 84%]

**4.2.b: hearing difficulties** [Agreement: 80%]

**4.2.c: consistency between siblings' language development**

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	4%	14%	16%	37%	30%	5 = strongly agree

**4.2.d: issues related to trauma, attachment, prematurity**

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	4%	9%	18%	36%	33%	4 = agree

**Statement 5:** The questionnaire should not aim to measure the child's language proficiency. This should be done by other means.

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	4%	22%	14%	34%	26%	2 = disagree

**Statement 6:** The child's proficiency should be documented for the language(s) that cannot be tested directly.

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	1%	8%	21%	45%	25%	4 = agree

**Statement 7:** The questionnaire should ask about difficulties the child may have (had) with language, in order to identify what might require further assessment by specialists.

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	2%	11%	13%	30%	44%	4 = agree

**Statement 8:** Exposure and use should be measured (for each language):

**8.a: over an average week** [Agreement: 85%]

**8.b: over the last 24 hours** [Agreement: 32%]

**8.c: over holiday and school periods separately**

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	1%	13%	20%	34%	33%	4 = agree

**8.d: over home and school separately** [Agreement: 92%]

**Statement 11:** The amount of overheard speech (between parents) needs to be estimated.

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	2%	7%	22%	44%	26%	4 = agree

**Statement 15:** Language history should be documented through year-by-year measures of language exposure and use.

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	1%	23%	18%	35%	23%	3 = I do not know

**Statement 17.1:** Time spent in school should be quantified. [Agreement: 89%]

**Statement 17.2:** This should be done in:

**17.2.a: hours per week** [Agreement: 78%]

**17.2.b: weeks per year**

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	4%	18%	15%	36%	27%	4 = agree

**Statement 26:** The following aspects are indicative of input quality:

**26.a: parental education** [Agreement: 78%]

**26.b: interlocutor proficiency in each language** [Agreement: 91%]

**26.c: interlocutor accent in each language** [Agreement: 34%]

**26.d: language mixing** [Agreement: 52%]

**26.e: pre-literacy and literacy activities** [Agreement: 82%]

**26.f: digital exposure**

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	1%	10%	27%	41%	21%	4 = agree

**26.g: playing with peers** [Agreement: 82%]

**Statement 27:** The language proficiency of the child's interlocutors should be documented (based on the respondent's estimate). This should include estimates for:

**27.a: each parent** [Agreement: 90%]

**27.b: any siblings** [Agreement: 88%]

**27.c: other members of the households** [Agreement: 77%]

**27.d: playmates**

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	2%	12%	17%	38%	30%	4 = agree

**27.e: teachers** [Agreement: 75%]

**Statement 28:** The language proficiency of the child's interlocutors should be estimated in relation to a native speaker of that language.

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage	4%	16%	20%	43%	18%	4 = agree

<b>per scale point</b>						
------------------------	--	--	--	--	--	--

**Statement 30:** Language mixing (in interactions with the child) should be documented for each interlocutor.

<b>Scale</b>	<b>1</b> Strongly disagree	<b>2</b> Disagree	<b>3</b> I don't know	<b>4</b> Agree	<b>5</b> Strongly agree	<b>Your choice was</b>
<b>Overall percentage per scale point</b>	3%	12%	13%	48%	24%	4 = agree

**Statement 31.1:** Language mixing should be estimated (in terms of exposure and use).

<b>Scale</b>	<b>1</b> Strongly disagree	<b>2</b> Disagree	<b>3</b> I don't know	<b>4</b> Agree	<b>5</b> Strongly agree	<b>Your choice was</b>
<b>Overall percentage per scale point</b>	3%	11%	18%	46%	22%	4 = agree

**Statement 31.2:** This should be done through examples (of the different types of mixing) rather than overall estimates from the respondent.

<b>Scale</b>	<b>1</b> Strongly disagree	<b>2</b> Disagree	<b>3</b> I don't know	<b>4</b> Agree	<b>5</b> Strongly agree	<b>Your choice was</b>
<b>Overall percentage per scale point</b>	4%	12%	29%	34%	22%	4 = agree

**Statement 32:** Language mixing should be documented separately for language use and language exposure.

<b>Scale</b>	<b>1</b> Strongly disagree	<b>2</b> Disagree	<b>3</b> I don't know	<b>4</b> Agree	<b>5</b> Strongly agree	<b>Your choice was</b>
<b>Overall percentage</b>	4%	12%	18%	46%	19%	4 = agree

per scale point						
-----------------	--	--	--	--	--	--

**Statement 36.1:** The literacy practices of the parents need to be documented (e.g. print exposure, reading the news, reading novels, etc.).

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	3%	22%	16%	37%	21%	3 = I do not know

**Statement 36.2:** This needs to be done independently of parental education and socioeconomic status.

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	2%	13%	26%	29%	29%	3 = I do not know

**Statement 36.3:** Parental print exposure is an informative measure of literacy practices.

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	2%	6%	36%	43%	13%	4 = agree

**Statement 38:** If a parent was educated in more than one language, this should be documented separately for each language.

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage	1%	15%	15%	41%	28%	4 = agree



per scale point						
-----------------	--	--	--	--	--	--

**Statement 41.1:** There should be a question about the status of each of the child's languages

**41.1.a: within the local community (including school)** [Agreement: 79%]

**41.1.b: in the "home" country (if applicable)**

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	1%	11%	15%	30%	42%	4 = agree

**Statement 42:** There should be a question on attitudes to language mixing

**42.a: within the family (at home)**

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	1%	14%	14%	36%	35%	4 = agree

**42.b: within the local community (including school)**

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
Overall percentage per scale point	2%	17%	18%	32%	31%	3 = I do not know

**42.c: within the broader society**

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
-------	------------------------	---------------	-------------------	------------	---------------------	-----------------

<b>Overall percentage per scale point</b>	2%	18%	26%	26%	28%	3 = I do not know
---	----	-----	-----	-----	-----	-------------------

**Statement 43.1:** Parents should be asked if they feel pressurised to speak the societal/majority language.

<b>Scale</b>	<b>1</b> Strongly disagree	<b>2</b> Disagree	<b>3</b> I don't know	<b>4</b> Agree	<b>5</b> Strongly agree	<b>Your choice was</b>
<b>Overall percentage per scale point</b>	1%	12%	13%	39%	35%	3 = I do not know

**Statement 43.2:** To identify the source of that pressure (if any), a list should be provided for the parent to choose from.

<b>Scale</b>	<b>1</b> Strongly disagree	<b>2</b> Disagree	<b>3</b> I don't know	<b>4</b> Agree	<b>5</b> Strongly agree	<b>Your choice was</b>
<b>Overall percentage per scale point</b>	2%	13%	22%	39%	24%	3 = I do not know

**Statement 49.1:** The questionnaire should be available in different lengths.

<b>Scale</b>	<b>1</b> Strongly disagree	<b>2</b> Disagree	<b>3</b> I don't know	<b>4</b> Agree	<b>5</b> Strongly agree	<b>Your choice was</b>
<b>Overall percentage per scale point</b>	1%	7%	17%	35%	40%	4 = agree

## **Supplement 4**

**Paper title:** How to Quantify Bilingual Experience? Findings from a Delphi Consensus Survey

**Content:** Compiled round 1 comments sent to the panellists with an invitation to complete round 2 of the online survey.

**Quantifying Bilingual Experience (Q-BEx) Delphi Consensus Study**  
**Round 1 Report: Comments**

**Statement 4.1:** The questionnaire should contain a section to identify children who might be at risk for a Developmental Language Disorder.

**Statement 4.2:** This should include:

**4.2.a: early language milestones**

**4.2.b: hearing difficulties**

**4.2.c: consistency between siblings' language development**

**4.2.d: issues related to trauma, attachment, prematurity**

<b>COMMENTS:</b>
<ul style="list-style-type: none"><li>• Children who are refugees, need support in all these areas; conversing with family members could help them.</li><li>• It is about bilingualism, not about language difficulties....</li><li>• see my comment above</li><li>• Evolution in milestones according to changes in personnel situation</li><li>• I guess with "trauma" you also have adopted children in mind?</li><li>• Hearing difficulties are important.</li><li>• Family history should extend beyond comparison with siblings to include parents, aunts and uncles, and grandparents. It should also include reports of concern on the part of adults outside the immediate family, such as other relatives, care givers, neighbors, and medical professionals.</li><li>• I assume this question is related to the previous point about being at risk for a DLD?</li><li>• I have given a lower score (2) to (c) because consistency between siblings' lang development (i.e. the relation of one child's development to that of his/her siblings) is rather tricky to interpret. DLD often runs in families - all siblings might be slow in development so consistency need not tell us much. For a different reason, I have given a lower score (3) to (d) as these issues might be touchy to ask about. I would ask about (a) and (b), as they are more 'absolute' in my view.</li><li>• I think D is an important consideration, but I think obtaining the data in this context may not yield the most reliable results</li></ul>

- - Siblings may have qualitatively different experiences so may result in unfavourable comparisons? - With the caveat that the survey users can, or be assisted to, make the distinction between typically developing bilingual children and those with additional needs
- As to question (d), we refrained from including such questions from our latest project questionnaire for refugees. We decided to refrain from such questions because we are linguists, not psychologists or social workers. We feared that these questions might cause anxieties. On the other hand, asked and interpreted by professionals, information about traumatic experiences is essential.
- I'm uncertain how hearing difficulties other than reported by parents, particularly for younger ages, would be identified. In my organization all children under the age of 5yrs who are referred for SLP services either receive a full audiology assessment or a hearing screener. Re: siblings- differences in language development/exposure are different enough that in my clinical experience comparing to siblings does not add value. Re: trauma: I am unsure some of these questions feel more like a full case history that would be accomplished from a full SLP assessment and may not necessarily be needed to determine whether a referral for a full SLP assessment is necessary or not. It would depend on who is using this questionnaire/screener and in what context that some of these more sensitive questions (ie siblings/trauma) would be appropriate.
- These things can be part of your case history
- I don't think this is our role unless we are conducting a speech/hearing clinical study.
- The questionnaire should focus more on language experience than developmental questions, although these are an overall part of interpreting the data and would pinpoint those needing further assessment/referral
- These are complex areas and will change the purpose of the tool. I think it will be better to have a tool that does one thing really well rather than tries to do several things badly.
- Not sure if this question refers to DLD or bilingual exposure
- Asking about early language milestones and hearing difficulties should be enough.
- As stated earlier, I don't see this being part of measuring exposure/bilingual experience. If this is not supposed to be a questionnaire or set of measures about language development. Naturally questions about language exposure through siblings need to be asked and conclusions about differences between siblings can be made based on that if needed.
- I strongly agree in the fact that these factors might impact on language development. I also agree on taking into account whether the participant being measured is a person with language or speech disorders. However, it is not clear to me how this information might varyate bilingual language exposure. From my understanding twins (one having language

disorders and the other with normal development) would receive exact amount of bilingual exposure unless adults vary it. Maybe the correct question to be answered in the questionnaire is whether the use of languages (at home, class) has varied with respect to other children because of their condition of language disorder.

- Plus parental concerns as a sensitive marker of DLD
- I am surprised to see hearing difficulties, trauma, attachment, prematurity related to DLD. DLD (previously: SLI) is what happens when NONE of these factors can explain a problem with morphosyntax. So you should strike the term DLD in Q5 and just write: language learning problems.
- as long as these are well established in each of the languages
- But most importantly, it should include information about the L1/ Heritage Language of bilinguals, which cannot be assessed by clinicians in many cases.
- it should also include if the child has other disorders
- having too much here may make difficult for parents to interpret/complete the questionnaire. In addition, usually, parents have different judgments for their children, usually related to order of birth
- Quantification of language exposure, use and exposure needs to be differentiated from medical case history - I do not recommend to try developing a one-size-fits-all tool, which then would not be recommended to use / accepted for data analysis e.g. by physicians
- I am not clear on what this is referring to. Is the goal to include information to provide to parents, or is this a separate set of questions to identify children at risk for language disorders?
- It should include other medical conditions and social communication problems that could have affected language development
- Siblings do not always have the same experiences, so that would need to be explored, too, in the case of discrepancies.
- Also include AAC, distinguishing between children who acquired language through AAC, and children who acquired language conventionally and currently use AAC. Conventional AAC may be in a second language and not a first language or vice versa
- I think one question is needed to flag any concerns or history of language impairment. This way the questionnaire doesn't become too long.
- across different languages, developmental milestones differ. Comparison between siblings has relevance only where their linguistic exposure has been identical - this is rarely the case

- and family history of learning or speech/language delays
- Consistency between siblings; language development is not easy to map -- in India with low SES group this would not work. At best some narratives from parents might be useful information that is if they are able to identify such problems.
- I agreed with the a, b and c as this could serve as useful information to investigate further in a bilingual study, not to screen for language difficulties, as I argue above. regarding statement d, if one needs to investigate this, an interview is the suggested tool to use, to my opinion this cannot be assessed with a questionnaire.
- B and C aren't markers for DLD, they provide important information but belong in the general anamnesis and do not belong in this questionnaire, in my opinion. But you need them to interpret the outcomes of the questionnaire.
- Points (c) and (d) could be part of optional questions, part of the core questions.
- hearing difficulties cannot be measured by parent questionnaire, unless you mean asking if the child has been tested for hearing impairment
- re sibling language development. Do you mean the DLDs information of an older sibling? Then I strongly agree
- I think most professionals will use their own questionnaire besides the questionnaire for bilingual children. So I would only use questions that are related to the bi- or multilingual development
- and environmental issues too
- family risk (not only siblings) but also parents, and parental concerns as this has been found to be a predictor
- I'm conflicted about including "hearing difficulties" on this list. I've chosen "agree" because I do think it's important to recognize that most children with reduced hearing don't actually have true language disorders: they simply lack sufficient access, and as such that's an important diagnostic criterion. My hesitation comes from two sources: (1) some children with reduced hearing *also* have true language disorders (with or without comorbid diagnoses), and I don't want to exclude them from meriting a language disorder diagnosis when it is appropriate. (2) With respect to deaf or hard-of-hearing individuals whose prime language-learning years are now behind them (e.g. school age and onward), it may be appropriate to consider continued difficulty acquiring a late-L1 as a type of "acquired" language disorder: acquired in the sense that it was probably not present at birth, but a disorder in the sense that even though their mind might once have had the potential to support age-appropriate mastery of a language, that ability is now severely diminished. So if hearing difficulties are to be included on the list, there should be a

mechanism for considering the child's current age in order to interpret that information most appropriately.

- I think the latter point (d) is particularly relevant in many immigrant communities in inner cities.
- In my opinion this needs to be a separate piece of work and cannot be assessed just through a quantitative method- in my experience that can be quite dangerous- so you would need a careful set of training around this.
- It is not really clear to me what 'this' refers to in question 4.2 (the questionnaire as a whole or the section on DLD)
- This should be a separate questionnaire and not part of a bilingualism questionnaire. There are already tools/questionnaires developed for assessing risk for DLD.
- Whether or not audiology testing has been completed may be more useful info. than asking about hearing difficulties as most parents will say that their child hears fine (even when this is not the case!)
- questions automatically asked in an anamnestic interview, for speech therapists, but which could be indicated for another audience.
- Even though I think a) and c) are important I have ticked no. 2 because these two criteria have been used in the past to identify bilingual children as deficient in some way. These two criteria should be part of the child's language journey
- I would add (e) exposure to each language of course
- it could include the child's current verbal communication skills, motivation and abilities too in each of the languages the child speaks.
- The answers to these questions depend upon the questionnaire also containing a screening component.
- If this section were included, then it should include this sort of information - but the section should certainly be optional in my opinion.
- Please see my comment for the previous question.
- It's unclear to me how informative (c) and (d) are with respect to determining DLD.
- I think it would be important to know if the child is typically developing or not - but the worry is that this would mean moving away from the (potentially) main focus of the questionnaire (bilingual language exposure).
-



- all those should be available but not compulsory in order to make the questionnaire applicable for studies with very big cohorts where data collection needs to be fast and efficient
- I see these are different questions from the question of exposure and use. Siblings can differ because language exposure in a family changes over time. Clinical professional are better suited to take these things into consideration
- Consistency between siblings may not be a factor, since no two children are the same or develop in the same way, not even twins develop the same way
- These issues can be covered in a general questionnaire on speech and language development but not in a bilingual questionnaire. This will also be more time consuming given questionnaires on bilingualism are long and typically has a lot of key bilingual areas to cover in a given time.
- Perhaps visual difficulties should also be included since blindness or visual weakness has been correlated in the literature with autism. There are criticisms of inappropriate diagnoses due to this correlation that could be detected using this questionnaire.
- Not sure what is meant by C.
- Points b and d should be included in different studies
- Prematurity can be a risk factor (with low predictivity), attachment and trauma are not

**Statement 5:** The questionnaire should not aim to measure the child's language proficiency. This should be done by other means.

#### COMMENTS:

- Why not?
- Specialists (SLTs) are trained to do this in depth. Detailed analysis needs to be carried out by someone who understands how to do this
- Proficiency: Are you including communication and speech, in terms of social and educational functionality, or just linguistic competence? In case of inclusion > agree
- This should be another means of acquiring information and not the only one
- If it is about measuring experience, then it should not necessarily measure proficiency. That is a separate aim.
- Why not? If you also want to screen for LDL, you should have some measure of proficiency in each language.

- It should be made clear that other objective measures of proficiency are needed, but having parent ratings of proficiency can be a useful addition to objective measures.
- I don't think you can 'measure' lang proficiency in 13 different languages with one questionnaire. You could ask about the parents'/caregivers'/etc rough estimation of the child's lang skills in language A and B, in comprehension and production, but to me that is not an actual measure of language proficiency.
- It could be useful to have this data, but it would also need to be obtained from other sources as well
- proficiency is important - not sure that it matters if it is assessed separately or not
- If we assume that impairment in L1 = impairment in L2 and typical development in L1 should equal typical development in L2, then this measure shouldn't need to be included in this tool.
- Could it not ask for the proficiency (albeit measured by other means)? This could be standardised by using CEFR for example.
- Experience and proficiency make be valuable to consider together
- The questionnaire should provide the background variables for the interpretation of language measures.
- generally I find that for preschool aged children (0-5yrs) a combination of parent report but also a robust Natural Language Sampling combined with formal assessment is necessary for accurately measuring a child's actual language proficiency.
- Again, as bilingualism researchers this is not our aim. Proficiency can be assessed via other methods.
- Parents could give an estimate of the child's proficiency in each language as this has been shown to be an important diagnostic indicator
- language proficiency should be content based rather than survey based
- We need this score to include into our statistical model
- How can childrens language be measured without considering their proficiency
- A questionnaire can be used to tap into proficiency but the questions might differ (just as questions around DLD might differ)
- As stated before, this should be about measuring language experience. In this way it would be most useful for several different situations and purposes.

- Do you mean in L1 or L2 or both?
- It is useful to get parental assessments of children's proficiency in ALL languages. Often there is no alternative.
- I disagree. In many cases there are no tools to assess the L1/ Heritage language of bilinguals and the questionnaire is the only window into the linguistic abilities of a child. As for L2, which is always assessed by the clinicians, the questionnaire L2 index can reflect the reliability / validity measure provided by a parent.
- The questionnaire could ask parents to rate their child's language proficiency in their home language as objective as possible. Eg. by having them answer yes/no questions about which communicative skills their child shows. If parents lack skills in the other language(s) the child speaks, their rating of their child's abilities might be unreliable.
- I think that it is likely that this would not replace formal/informal language assessments including observations, this would still be needed but it would be useful to have some indication of language proficiency.
- at least as an overall measure, it's a good idea to have this
- Measuring language proficiency can of course be done by other means, but I have no objection to including some questions that ask parents to rate their child's proficiency in each language.
- the most important aspect of it is to identify whether the language development is typical or atypical (especially that there are no clear milestones for bilingual language development).
- I would use measures that are already established for proficiency
- I think in an ideal scenario the questionnaire would also measure children's language skills, either through relatively objective measures similar to those in the CDI report forms, or through more global statements (as in ALDeQ)
- This means as well as complementary tools like language diary, some tests may also be used. In Indian primary education context standardized tests may not work as the tests have a western bias (understandably so!) and are far too expensive for us. Even our university will not purchase for us meaning to do good research. So local tests have to be designed, used, documented and validated.
- The tool should however allow one to identify if there is a problem and need for more in-depth assessment.
- This can be achieved by a questionnaire (ask how well the person reads, writes, understands and speaks the languages used). Additional measures could be used (such as

a Confrontation naming test, Vocabulary test etc.), however these would serve only as additional tests that can further validate the assessment from the questionnaire. That is, I believe that the aforementioned questions on perceived language proficiency (asked in both the parents and the children themselves) is a rather good indicator of language proficiency. Given the subjectivity of such questions though, I also suggest to use an objective, computerized task to assess balance between languages used (such as the one used in my studies e.g. Ladas, Carroll & Vivas, 2015)

- It provides information about the level of interaction in that particular language
- The answers in the questionnaire may support a professional's assessment of language proficiency.
- Parent assessment of proficiency in the L1 in particular could be helpful if it is unlikely that the L1 can be tested because it is an understudied language
- This depends on the age of the child. Parent report may be the most reliable measure of language proficiency for very young children. Even if children's proficiency can be measured directly, this is still valuable information to gather in a questionnaire.
- I think it is essential to differentiate between quantifying the input that is available to a multilingual learner and that learner's mastery of that input. I would be amenable to a questionnaire that aimed to satisfy both ends, provided that each is distinguished from the other. Conflating the two (or measuring only one) will not address major questions of both theoretical and applied interest.
- It should not "measure" the language proficiency but it could include some questions about it, especially for the language that cannot be tested.
- I am not sure if you mean first language proficiency?
- This has to be done through observation over time, not in a single setting. The Bell Foundation and Hounslow systems work well I think, so no need to repeat them.
- To evaluate real "proficiency" in a language, one would need to conduct a series of standardized language tests. But a parental estimate of proficiency could be useful.
- Most of the time, dominance can be determined through our assessments, observations, and parental interview. However, a few questions on this subject could be useful in cases where language use is similar between languages.
- but it is difficult to quantitatively assess the linguistic skills of a bilingual child...
- I find self-rated proficiency measures to highly depend on cultural and other social/personal biases.

- I take this to mean direct measures of proficiency, which is different from e.g. asking the parents (or teachers) their evaluation of proficiency.
- If by measuring proficiency, it is meant "parents' report/evaluation on the child's language abilities", then I do not think it would be reliable anyway, just like in any other case of self-reported/evaluated language proficiency scores. Adding other layers of more standardized measures of language proficiency could be impractical as most things will be language sensitive and language specific. So, I think this should be left out.
- Any measure of reported proficiency will inevitably be so minimal as to be likely be useless.
- I can see this is problematic though - it is hard to tease these apart in some ways
- Some amount of information regarding the language proficiency of the child is necessary. If there is a separate tool, then ways of integrating that with the questionnaire has to be arrived at.
- There are different assessments for expressive and receptive language and if these were included in the questionnaire it would probably be too long.
- It probably shouldn't (and cannot) measure proficiency directly; but perceived proficiency level reported by parents could be very useful, similarly to how it's done in BiLEC
- Proficiency should be measured with tools that have been developed for that and which need to be language specific. At the most, the tool could include the parents' rating of the child's proficiency
- Measuring language proficiency is important. While I understand that the term proficiency means different thing to different people (e.g., complimentary principle) and it is hard to define the term 'proficiency', there must be some attempts to capture this in some form. Avoiding this important aspect of bilingualism wouldn't solve problems associated with measuring proficiency.
- The questionnaire may contain an additional section to measure the child's language proficiency. In this case, it is necessary to define from which theoretical framework the language proficiency must be understood and measured.
- No need to measure, but a item stating the level of the student had better be included
- This is true, but often impossible for the home language. Are we only interested in the societal language?
- I agree that measurement would be very difficult but I believe you need to try and have an understanding of the child's proficiency in each of their languages, to understand their

context and begin to unpick any needs/gaps/concerns in any of their languages. An overview could be gathered from parents/carers or staff working with the child.

- There might be some items giving information about language proficiency, although such kind of measure is not the aim of the questionnaire

**Statement 6:** The child's proficiency should be documented for the language(s) that cannot be tested directly.

**COMMENTS:**

- We can't be sure that the proficiency recorded is accurate without an objective measure
- But who is capable for commenting proficiency?
- Why can't the language proficiency of a language not be tested?
- that is a good idea, to have a possibility to assess language if really there are no other ways
- Dynamic assessment can also be used to look at child's language learning potential
- Yes, also with a measure to report previous and actual use
- Maternal education may impact on how accurate reports of home language use are. Many parents living in poverty may compare their child with other siblings or those in the neighbourhood and these comparisons are false, since 50-70% of children living in poverty may have language difficulties due to poverty rather than being typical language levels.
- Yes, if you want to measure it, then it should be in all languages.
- I'm not clear on which proficiency measures you would use that would not be assessments.
- this begs the question of how
- We have other means to document language proficiency - but it's good to have multiple data points
- this is often not possible
- From where will this information be obtained?
- Not sure about this.

- Would these be tested later? Or would the proficiency be accepted without further inquiry?
- If this can be done by parents' or teachers'/caregiver's estimates, I would strongly agree.
- if the questionnaire should not assess the child's actual language proficiency I am uncertain about this statement.
- This statement is quite confusing.
- As above, parental concern and opinions are valid in this context
- it would have to be very clear where this information was coming from, how collected, how subjective etc doesn't it contradict the previous question a bit?
- What does it mean "cannot be tested directly"? Rare languages? Community languages? Underresourced languages?
- I'm not sure what this means
- I think proficiency and use should be measured separately (because they are different things and there should be objective measures of proficiency, if possible). I think this holds for all languages.
- This again should be done in other measure/questionnaire.
- This is problematic if proficiency would be documented ONLY for the language(s) that cannot be tested directly since comparison of proficiency tested directly (in one/some language/s) and reported in the questionnaire (in other language/s), would be very difficult. Thus including reported proficiency for ALL languages could be helpful.
- Well not only for those! For ALL languages in as far as parents are proficient in these themselves
- but how?
- The questionnaire could ask parents to rate their child's language proficiency in their home language as objective as possible. Eg. by having them answer yes/no questions about which communicative skills their child shows.
- this is ambiguous, who is going to provide this information when sometimes parents are unable to determine this aspect
- W/ref to No. 5 unclear: no measurement; only estimates, self-assessment should be documented

- I'm not sure this should be obligatory, but including questions related to language proficiency seems useful.
- I think I agree, but I am not sure how you would document a proficiency level that has not been established, e.g. by some form of testing or assessment. Does this mean taking reports from family members?
- particularly given the absence of validated measure for many languages
- I think assessing proficiency for the languages that can be tested is still worthwhile as not all children (especially very young children) can be tested, and testing is not always possible (for multiple reasons, including the current pandemic), more expensive etc.
- Such complementary methods of data gathering as I have mentioned above will be rich information to analyse and present.
- For all languages used, no matter whether these can be tested directly or not. But there should be separate total scores for perceived language proficiency for each language used.
- see comment for #5
- I am not sure if I understand this question.
- Do you mean that the proficiency of these language(s) will be asked in the questionnaire? And only the language which can be tested, should be measured? Is that the case when there is no translator available in one of the languages? Or do you only measure the language of the country you live in? In my opinion, I think it is important to know what to expect from all the languages and only measure the best language(s) of the child.
- This question makes me think. I think a further discussion is required.
- but different instrument
- It would be best to document child's proficiency for all of their languages (even when some of them can be tested directly), so that there is a comparable score across the languages.
- I'm not sure what is meant by "documented" in this case... I assume this refers to a case where the investigator lacks either the tools or the skills to conduct a formal language assessment. In that case, it seems better to at least obtain some sort of informal estimate of the child's proficiency in the non-assessed language, but certainly this information would need to be treated differently than its formally-evaluated counterpart.
- I would have thought this would have a bearing on whether the child has DLD or not.



- Nothing should be missed out. This is vital for children who have not had much prior formal education in their home country.
- This would be very helpful in the cases where interpreters are not available and where we question parents' judgments about proficiency.
- Should be documented for all of the child's languages
- How would it be documented if it cannot be tested? This is unclear.
- That's actually not a bad idea but again a clear reference level should be given. For example, language proficiency related to native speakers or other speakers in the community? This reference level could be clarified in the survey instructions or in the question itself.
- I take this to mean asking for parents' or teachers' evaluation - if possible, something more would be welcome.
- Again, I am not a big fan of self-reported proficiency measures. However, I think the child's experience in her/his other languages should also be documented as detailed as possible, which then could be used as a proxy for their overall proficiency in that specific language. Maybe this could also be included in the calculation of individual and/or cumulative scores.
- documented in the questionnaire I suppose?
- If this is possible in sufficient detail (what is sufficient is an empirical question) and as an additional add-on only. But perhaps it would be sufficient to refer to the ALDEQ for this. I think you need to be wary of unnecessarily re-inventing the wheel .
- How would that be possible for any language involved?
- Documentation of proficiency levels of different languages should be uniform.
- Not sure what is meant by “documented” and who is doing the documenting
- This seems to address a different question than exposure or use. It is a complicated issue that I don't see as the focus of this project
- Question number says we are not to focus on language proficiency, why do we now have to check proficiency for languages that are not going to be assessed directly? Maybe I'm not understanding the 2 questions
- The child's proficiency in each language should be evaluated in the same way.
- Proficiency in all languages should be documented if at all possible.

**Statement 7:** The questionnaire should ask about difficulties the child may have (had) with language, in order to identify what might require further assessment by specialists.

**COMMENTS:**

- And this should be in all the languages spoken by the child
- Language related difficulties needs to be specified, eg. problems with forming sentences, usage of verbs etc.
- I'm answering here on the basis of what I said to question 6: there should be a part of the questionnaire which assesses language if there are no tools in that language, and in that case of course questions about language difficulties are important.
- Yes, for example inflection and agreement in verbs in rich morphological languages or prepositions and adverbs for other languages
- This should be done by asking parents to compare the child to other children in the same situation or in the home country (if possible)
- again, would this aspect be linked to the screening portion of the questionnaire? I would want each to be separate in order for the purpose of each question section to be clear to the administrator as well as the parent/guardian answering the questions.
- I think this sort of questions would make the questionnaire very long and inappropriate for children or busy parents.
- yes, it will be very informative on the part of the project team. By eliciting the problems from the child, it becomes all the more interesting to predict. May be should document via recording too
- As part of any language assessment I would want to know what parent/carer/education practioners views of the child's difficulties were so for me this would be part of regular case history for any child. Is the questionnaire hoping to identify children that might need a language assessment or is it part of a language assessent giving information to help reach a diagnosis? or both?
- Is it going to be possible to operationally define "difficulties with language"?
- Again, this seems like a slight different purpose
- The questionnaire should consider for example attitudes towards language use and hearing it exposing oneself in the situations where different languages can be heard and used, but language difficulties shouldn't be the scope of this questionnaire.
- How this differ from the question about potential risk for DLD?

- ask whom? teacher? parent? they may not have the experience
- Yes, if it is used for education or therapy purposes. This could be a separate section.
- From my experience, parents often don't know the answer to this question, so examples of difficulties may be needed, such as difficulties understanding instructions, difficulties using sentences etc
- but too many questions about this may deviate the purpose of the questionnaire
- Including AAC users and making allowances for children who have acquired their language through alternative or augmentative means has to be considered.
- a definition of 'difficulties' is necessary
- you need to differentiate between children who have difficulty with acquisition of any language in the presence/ absence of appropriate stimulation, and the child who has had difficulty with acq of a second language in the presence of appropriate stimulation
- Perhaps, this question as well as the previous ones about developmental delays/disorders could go into a separate section that can - but need not be- administered. So, the core part would be about quantifying children's language exposure and use, and in clinical settings or in other situations where information about potential disorders/delays is needed, these additional questions can be administered?
- Children's self assessment and sharing of such information would be a good information source apart from parents/care givers.
- Bu these should only be screening questions, down to a minimum to avoid high drop-out rates
- This needs to be restricted to the language(s) parents can make this judgment about. "Difficulties" can be vague and I think a place for a qualitative answer should be included here
- I ask parents of monolingual children those questions aswell, so it is already in my standard questionnaire. When I see a a multilingual child, I ask these questions for all the languages. So I think it is important to ask it, but I don't know if it is needed in a questionnaire for multilingual development. Because of the additional questions you need to ask partents of multilingual children in little time, I would prefer a brief, focused questionnaire with questions you don't ask in a monolingual questionnaire.
- I think not only difficulties should be asked, also strenghts.
- This should include a list of options from which the caregivers can select the area(s) of difficulty instead of having an open-ended question. In that way, data can be more comparable across studies.

- My agreement is contingent on the questionnaire having separate sub-sections that measure the input in the environment versus the child's mastery of it. These questions would naturally go into the latter.
- Especially relevant in research to define exclusionary criteria
- Possibly, this would be better than having a separate section within the questionnaire on language delay- it would enable concerns to be flagged up and then taken up by professionals separately. I don't think the 2 should be mixed.
- May not be necessary as we typically ask these questions in our initial interviews.
- I think that if the purpose is to quantify the bilingual experience this is the only goal. Identifying difficulties may be an optional part of it
- I'm not sure about this one - the questionnaire should not be about 'diagnosing' language difficulties
- Adding examples for the various language domains for parents to better understand the components of languages (ex. vocabulary, narration, comprehension, reading/writing, etc.)
- Agreed but for this to work, it needs to present different contexts of communication to the parent so that s/he will be able to link language difficulties with specific situations the child experiences.
- This is only in case the questionnaire DOES serve as a screening tool.
- Again, this sounds very useful but might run the risk of causing unnecessary "worry" on the parents side. What could be more useful is to decide at the very beginning whether this questionnaire will be used with only typically developing children or not. If the answer is yes, then some standardised questions could be added to the beginning as part of inclusion/exclusion criteria.
- I doubt many of the people using the questionnaire will have enough knowledge to put such information to adequate use. This is a job for speech language therapists.
- This depends on what this questionnaire is for. If it is for diagnosis of DLD then yes. If it is for measuring bilingual language experience (only) then no
- It might be trying to achieve too much with one instrument. Maybe the instrument can have an optional extension that can be used for identification of at-risk children?
- I see that as a different question. A tool to measure exposure and use for research is not a clinical screener and should not be. The literature on what constitutes risk factors is

complicated and evolving. I think it should be left out of this. This is complicated enough already

- Again, beyond the scope.
- I am not sure this will benefit the objective of the study
- This is important not only in order to know if there is need for further assessment but also in order to understand how to interpret the other findings re exposure/experience/language. We can always identify children as outliers after data has been collected and exclude them, but knowing a priori what to expect and who might be a potential outlier is definitely better practice.

**Statement 8:** Exposure and use should be measured (for each language):

**8.a: over an average week**

**8.b: over the last 24 hours**

**8.c: over holiday and school periods separately**

**8.d: over home and school separately**

**COMMENTS:**

- with the child being exposed different languages; different stages of their lives will differ in how their languages develop
- Definitely.
- Depends on personal or group situation
- I would also include a month period of time as many international families have situations where one parent is absent for 2 weeks in a row or more, and where the clearer insight into the language use is given after 4-5 weeks...
- Children of all ages spend longer at home than at school. However, parents may deny home language use, so this also requires some explanation, since parents may feel that professionals want to hear that they only speak the language of school at home for educational success motivations.
- That's comprehensive!
- this would be difficult to measure over any amount of time
- I find this question hard to answer. I do not know of studies that have investigated (on a large scale, in different settings, for different ages, with different types of parents etc)

whether (a) or (b) or(c) are generally more reliable (e.g. whether (a) or (b) might be enough or not, etc), though this would be good to know. I have rated (d) the highest because it is often asked and in my experience easy to answer by parents and less work than filling in questions for every day of the week - of course it may also be a rougher measure.

- for d: it gives the professional information to know where the child is exposed to and uses which language.
- Over an average day. I don't see the need of evaluating holiday periods
- A) May be difficult for the parents to recall accurately B) May be easy for the parents to recall accurately, but may not be representative of a week C) Again, potential difficulty with accuracy
- The bigger the sample the better - 24 hours may not yield typical language use and exposure. Potentially too many confounding variables in that measure.
- Since groups of heritage speakers differ in schools and classrooms, especially (d) is important. A strong heritage group in a class will certainly change the proportions of use and input of the heritage and the majority language.
- Gathering information about different life contexts for the child can be valuable
- differentiation should be made between a typical weekday and weekend
- May be more challenging to gather information that is more remote but clearly this is important in many bilingual contexts
- Average week should preferably be not taken from one week only but spread in several weeks. Home and school/kindergarten should be separated and taken into account both. Studies that have referred to possibility that kindergarten environment wouldn't bring any extra in terms of language exposure cannot be generalised to every society. However, it need to be strictly explored in which kind of setting the learning is happening: what is the group size and activities to do with language and which languages the children and teachers use. Both kindergarten and schools need to be taken into consideration. The concept of those two are also different in different countries.
- (b) may be more of a challenge for parents and/or researchers to be able to document.
- Currently we record exposure to L2 and L1 on a separate cumulative basis but I have considered the question of (relative) exposure to L2 in non academic/school periods of time.
- Possibly other sources of exposure/use except for home and school should be included/ possible to report?

- All of it and more
- I think options a c and d would be more reliable in giving an accurate picture.
- unclear: measured clear: different settings should be acknowledged differently
- It is important that the questionnaire targets mostly the most relevant information so it does not become too long and therefore hard to use in a clinical setup.
- options c) and d) do not seem as clearly different as a) and b). I would think that home and school usage might well be different and should be measured and recorded separately. But if we take a short period, like a week, then it would be better to take a typical week, during school time, when we could see the usual mix and proportion of languages used.
- however I do not feel you should be calculating combined scores for the above
- This model might not work for the low SES group in India. But for medium and High SES this kind of fine tuned distinction of language use at several points of time in a year would be useful.
- And with whom (with parents/siblings, with other relatives, with teacher/friends at school, with friends at leisure time). This allows us to roughly estimate the amount each language is used.
- For d) it might be best to use 'out of school' instead of 'home'.
- On a typical period; this could be a typical week in the last year. Tool can ask if typical weeks are common or not in the child's life. If in unusual cases, a child let's say spend a part-time holidaying and part-time schooling then guidance in the tool could be added to the interviewer and respondent.
- I have put disagree because I think it is unrealistic to ask caregivers for such an amount of detail
- For children who are younger the school age (and for measuring cumulative experience during those periods), childcare environments must also be considered.
- Making sure that home exposure is valued is vital and this needs to be done over time. The holiday period is crucial to capture.
- I hesitated about b) but this detailed analysis would generate data of daily and cross-linguistic use
- It could be over an average "typical" week, excluding holidays.
- I think all these are relevant but not to the same degree for everyone.

- I would say rather a) than b), but choose 1. d) definitely!
- Again, it's an empirical question, but I suspect that (b) would lead to considerably large margins of error.
- I believe from my experience a lot of children are being robbed of the opportunity and benefits of being competent bilinguals due to inadequate and inconsistent and limited exposure of the home language.
- Ideally, over home and school separately and should include holidays as well. But, in reality it might be possible to monitor only over the average week and with a little effort in both the home and school contexts.
- While it would be good to see the difference in language ability and usage during school and holiday periods separately, I think it would take too long to do this if you have to wait a few months for a school/holiday period.
- I would agree with all four ways of measuring exposure. I think exposure at school and home are critical and should be measured separately. In many communities, language environment and exposure changes so quickly due to increased free movement between people (e.g., people living in a big metro or inter state borders) - measuring language exposure over the last 24 hours and over an average week becomes crucial.
- The frequency of the evaluation may depend on the age of the child. In younger children, the evaluation frequency can be daily, and for older children, it can be weekly or even monthly.
- d) I suggest incorporating the possibility of adding other contexts in cases where needed (for example, the social environment, in cases where the school language and the family language(s) do not coincide with those of the environment). Additionally, in the home context, I recommend separating the experience with each of the child's reference adults.

**Statement 11:** The amount of overheard speech (between parents) needs to be estimated.

**COMMENTS:**

- and parents to children... (Not only speech but also cultures appreciation)
- Okay you included this :)
- The role of overheard speech is not yet well understood. But measuring it could clarify.
- Amount of overheard speech between adults in the home (not exclusively parents).
- I don't know. I wanted to put "2", but perhaps in some contexts/ages this is more important than I can imagine?



- Would be good to know but question how reliable it would be
- ideally yes but this is very difficult to estimate
- I don't know what the impact on language proficiency of the child will be for overheard speech
- Not ideal, but tricky to work around if not estimated.
- If parents are only speaking a certain language when directly to child, but otherwise speaking another to each other, in community, with siblings- this greatly decreases the percentage of actual language the child is hearing. (ie French directly to child during morning/evening hours, English to each other, siblings and in community)
- I will be interested to know
- This is often lacking and it might be the third language for example. In the final calculation its importance might be however weighted in another way compared to direct contact with the child.
- I would be keen to learn more from others' experience about the impact of this on bilingualism/language acquisition. I certainly encourage parents to continue to use L1 even if their children are not using it so the children's exposure is maintained.
- Gosh, how accurate parents may be in estimating this? But I do agree that this may form an important source of lang exposure. Especially in some cultures!
- nor sure this is possible
- It is important to document the language used by parents, not only for the quantity of input. It might be an important factor in shaping the family language policy.
- to add: needs to acknowledge any overhead communication (see comment on patchwork families, child custody, siblings, ...)
- Though this would be hard to do accurately, in reality!
- This is important since children can acquire at least passive competence in overheard languages; and the linguistic and cognitive effects of passive exposure are becoming more evident in research
- I think this could have important implications.
- needs to be better than an estimate
- I'm not sure that parents will be able to estimate this.

- Overheard speech such as this has a tremendous influence on the child's language system!
- = What language the parents speak to each in that the child might overhear (assuming they might use a different language when the child might not overhead) ?
- Difficult to estimate? Biased responses from parents?
- Depends on whether there are limitations in terms of length of the questionnaire.
- Not sure how this would be estimated
- Perhaps just documenting which language parents use most often when communicating with each other
- I'm not sure whether this question is asking about absolute or relative amounts. I would be disinclined to support measures of absolute input. Including this measure for relative input would only matter for parents who report speaking to each other with a different distribution of input than they use with the child. Perhaps this could be dynamically included only for those families for whom it is relevant?
- Agree, if the language used between parents is not the same as the language parents speak with child
- This could be hard in families where parents switch when talking to each other. Any measure of that that is taken by a questionnaire must be interpreted with a pinch of salt.
- But perhaps also between and among other family members an carers
- That would be an interesting measurement but hard to see how it could be calculated.
- The language spoken between parents should be documented, but amount of overheard speech seems very difficult to accurately assess.
- if this is also a language that is used directly to the child
- Whilst I can imagine this might have an impact, how on earth would you expect parents to be able to estimate this? I think this is really asking too much of parents. What I think it \*would\* be worth documenting, is \*which language(s)\* parents use between themselves.
- Gaining this data may not be possible on numerous grounds: 1. access to the homes may be problematic 2. availability of both parents or all care givers at the same time may not be possible 3. There will be unnatural speech in terms of both quantity as well as quality in the presence of an outsider.
- It will be difficult to know for sure so an estimate is probably best.

- The language spoken between parents is important. Is it the home language or a different language.
- Sometimes parents use different language to hide sensitive information from children thinking that it will be difficult for a child to grasp
- While I agree that this needs to be documented, it's unclear what is referred to as 'amount' here - it's also important to specify the language that's overheard, if possible!
- Why just between parents? Why not between other members of the households, e.g. between two older siblings? Where does one stop?

**Statement 15:** Language history should be documented through year-by-year measures of language exposure and use.

Scale	1 Strongly disagree	2 Disagree	3 I don't know	4 Agree	5 Strongly agree	Your choice was
<b>Overall percentage per scale point</b>	1%	23%	18%	35%	23%	

**COMMENTS:**

- More important are the milestones: when exposure to other language started, when changes, when kindergarten started, when school ..
- I don't think there is a need to be super precise about the year by year history
- Depending on the relevance to do so. As a clinician (SLT), I discuss language usage at every assessment point with the families and see change more and less frequently than a 12 month period depending on family and child circumstance - of course, for research purposes there would be very different necessity to do that
- Good to follow evolution, but the last months are the important to know in order to program the best practice possible for the child (not for the school...)
- Or month-by-month, for infants
- Yes, and according to any significant changes in a child's life/language circumstances.
- This might be possible for very young children and highly dedicated parents, but in my experience it is not feasible/reliable when children get older.

- ideally but very time-consuming - depends on purpose
- In general. Not year by year.
- In very young children (in whom yearly change is more vast), this may bury frequent or subtle changes in language exposure and use.
- If this can be done by rough measures such as "1-3 only home environment, 3 kindergarten entry, 5 half days a week etc. " this would be helpful to calculate cumulative exposure, it can, however, be too much for some families and, in some cases, it might be impossible (as in some refugee family constellations we have encountered).
- this is particularly relevant should the child begin a new childcare program in a different language.
- Yes, in an ideal world we would measure everything. But this is a lesser priority as it is quite long to administer.
- This would be excessive
- if the child keeps travelling from one linguistic place to another because of their parents' job, it will be interesting to see
- If it is possible
- Ideally
- If these measure can be trusted and caretakers feel confident in answering those questions.
- in example, by including information about the number of changes
- In L2 acquisition/history - yes.
- Depending on the child's age it may be quite difficult to obtain a year-by-year measures. For younger children it can be feasible, for older - I am not that sure.
- But one cannot expect highly detailed information for years that are long ago. Language history should not be measured the same way for all preceding years
- it's a great idea but is it practical
- Only when it concerns an ongoing therapy or schooling trajet for special needs
- Not for every interlocutor. Yet, year by year in order to derive the index of cumulative exposure.

- Age(s) at which language exposure and use changed should be documented.
- My only thought is how long this would take and also would parents be able to remember to answer this accurately.
- Year-by-year would make the questionnaire for older children too long to be practical. Maybe by developmental stages (e.g. pre-verbal <12 months; 1-2y, 2-3y in case changes due to free nursery hours for 2 year olds; 3-4 in case changes due to free nursery hours for 3 year olds; and then reception - end of KS 1; KS2 etc)?
- I am concerned that this would take too much time to be of significant benefit.
- this information linked to questions about "special" events could bring new information (parent's marriage, divorce, place of residency)
- instead of year-by-year it rather should depend on individual occasions that may lead to change in language use, change of interlocutors, schooling, ... (e.g. see refugees)
- This level of detail should not be required, though some studies may want it.
- In order to make it more relevant and handy, maybe check if there are phases of different language exposures and then measure the cumulative language exposure by phase.
- I think this is important, but perhaps not a priority for a measure that will be used as part of a battery... This will be trickier for parents to fill out if in questionnaire form and from a logistical perspective (and recent experience) its tricky to work out how much you can expect parents to do as part of a study (particularly if this is part of battery, as mentioned above).
- I think current and recent exposure and a general info about history should be enough
- change takes place much more quickly than within the space of a year
- The year-by-year measures should only be used if there was a big change - so in our context children are often exposed to and speak the parents home languages - and then later once they start school (at 5-6 years) would then be exposed to a third or fourth language. It may be better to investigate and ask around changes in language exposure rather than a year by year measure.
- Overly detailed and lengthy this would be...Alternatively, you can use questions that detect any changes in the child's language environment (e.g. travelling abroad, changing country of residence, a caregiver different than the typical one introduced in child's life etc.).
- This seems overkill in a situation where the family moved to another country. In this case, exposure was probably fairly consistent (although of course different) before and

after the move. So perhaps parents should be allowed to choose their own period for reporting? Or at least there should be an easy function built-in to say "copy data from year X to year Y" if it is going to be a year-by-year report.

- if possible, it would provide good insights into the multilingual development of children. I just think it is not feasible in practice. I do think it is necessary to follow a child's language development over more measuring moments, but not for the purpose of documenting history of language exposure and use.

- I am not sure what the alternative would be and can imagine that this would require some brief comments/explanation to be able to interpret any changes.

- too detailed for anyone to fill in

- This seems cumbersome for parents and difficult to do reliably for parents.

- This could become tedious for older participants. Perhaps a procedure could be established such that for children under age 3 (for example), the interval is smaller (maybe 6m), then grows larger (maybe 1 year) for children ages 3-12, and then even larger for ages 12+ (e.g. 0-3, 3-5, 5-12, 12+)? Weighting those periods appropriately would be a challenge, though...

- It would be more precise but become really time-consuming for older kids.

- I think this might be too difficult to gather this information.

- Possibly useful to see when a child started to learn a majority language, not sure it would need to be year by year.

- it depends on the child's age. For preschool children we should, but for school age children, language history from preschool years may not be easy to obtain year-by-year especially if the child has already been in school for 3 or more years.

- I think the questionnaire should be able to capture milestones in the child's experience (e.g. nursery, primary school etc.). A year-by-year measure of granularity may not be necessary.

- I only disagree as I am not sure it would be practical and/or reliable as it might not be easy for parents to be as specific as we desire them to be when answering such detailed questions.

- what are we going to use this for?

- Maybe use age ranges (e.g b/w 0-3)

- Whilst I applaud the idea behind this, it is rather hard to do. I think that it's easier for parents of younger children and for parents who are more aware of their bilingual

parenting strategy. Perhaps this could be an add-on for those wishing to use this because they think it is appropriate for their target group and who are certain the parents will be able to fill it in.

- Yes but I guess this would depend on the age (if someone wanted to use the questionnaire in this way). For v young children a year is too long as language development happens more rapidly - for older children year by year might be OK.
- Albeit time consuming, this will be a very interesting data.
- this method does permit a look at changes over time
- This action may require an additional instrument to monitor each participant. Informed consent for research may vary for each country. Informed consent models must be included as part of this instrument.
- I think that it is enough to define the exposure and use by periods in the child's life, regardless of its duration (as long as they are periods of more than a month)
- Good in principle but increasingly difficult with age of child
- If we are going to track the student's language use everywhere as mentioned before there is no need for that documentation. Milestones can be tracked according to each student's journey. I think this would provide too much info that can cause confusion during analysis
- Depending on the child's age this can only be possible in very general terms, e.g. move to another country, beginning of childcare or school, birth of a sibling.
- Yes for year-by-year, but it does not mean that the same level of details should be applied for every year.
- Ideally, but very hard to do in practice.

**Statement 17.1:** Time spent in school should be quantified.

**Statement 17.2:** This should be done in:

**17.2.a: hours per week**

**17.2.b: weeks per year**

<b>COMMENTS:</b>
<ul style="list-style-type: none"><li>• Or at least it should be stated whether the time in school is regular or irregular</li></ul>

- They might absent for 2 weeks due religious reasons; like now the schools might be closed.
- It is difficult to quantify weeks per years for parents
- Both parameters are useful - the child's attendance should also be considered as this is incredibly variable and, particularly with younger children, is sometimes based on cultural expectations
- Proportions of activities could be better: class, special needs support, sport, play
- And adapted to changes throughout the year – sometimes depending on changes in curricula...
- Terms.
- I don't know whether parents off the cuff know weeks by year. If it is known how long school terms are, maybe (b) is not necessary? instead you might want to ask about absences (e.g. from preschool)?
- both
- Rough percentage
- Both.
- In refugee contexts, an estimate for (b) might be more informative
- only if more than 6 weeks approx
- It probably doesn't matter if the calculations is then taking into consideration the school holidays. Maybe separate question about hours per week and weeks/months of school holidays would be easiest for caregivers to answer. Then school hours "missed" during the holidays could be easily calculated.
- Certainly for (a) above in context of part time/full time Nursery.
- The data obtained from the questionnaire could be later compared to the independently obtained statistics for each country/region (if that exists!).
- Combination of both measures will yield a good measure that will allow comparisons amongst children even though one will not know the exact amount of input in the school language, given that we have no measure of input frequency at school
- I'm not sure this is such an important measure
- I would say hour per week. The hours that children spent at school vary greatly.



- Important to make measures internationally comparable
- the information regarding weeks would be useful if the child has a significant period of time when he does not attend school
- We need both to get an accurate picture.
- both combined
- weeks per year only if you have indications for different amount of weeks per year from some countries, as this is rather typical and similar for most countries
- See comments in previous question
- see comment for 17
- too detailed
- An ideal instrument would accomplish this largely automatically, allowing the user to simply enter a parameter for each of these values. Many families may not know; for these cases, the investigator may benefit from guidance about the usual practice in the family's area and use that as a default.
- Surely you need both?
- I think both might be necessary for different context. For the mainstream school, weeks per year would be fine but if there will be a subsection for extra-curricular activities (e.g., weekend schools), then hours per week would be more meaningful.
- I suspect that hours per week is more informative but this is little more than a hunch. Perhaps you could take a sample of countries and determine whether one is subject to more variation than the other?
- I think especially for young children who are just starting school and where school language is different this will be important.
- I do not think there is much variability in weeks per year.
- These are different measures that complement each other.
- Different countries, and even different regions or school systems, have very different numbers of hours of school per week.

**Statement 26:** The following aspects are indicative of input quality:

**26.a: parental education**

**26.b: interlocutor proficiency in each language**

**26.c: interlocutor accent in each language**

**26.d: language mixing**

**26.e: pre-literacy and literacy activities**

**26.f: digital exposure**

**26.g: playing with peers**

<b>COMMENTS:</b>
<ul style="list-style-type: none"><li>• Digital exposure depends on the type and medium of exposure e.g. how interactive and what are they getting alongside that - are they playing on a tablet alone or with a parent etc</li><li>• Important for school training, not so much for communication at home, street and work</li><li>• The interlocutor proficiency in each language and the accent are not necessarily indicators for input quality: many non-native speakers provide a greater and more consistent input than others, are more aware of the language they use and how they use it than some native speakers. The code-mixing is a common practice in multilingual families and, like translanguaging in class, can help (or not, if not done efficiently) Although digital exposure can help, in some cases it can be rather hindering (when the digital aid is used in the wrong way)</li><li>• Language mixing/code switching would be regarded as normal and is very frequent. Most parents would not know if they code switched. It is largely irrelevant as if information/content is inserted then home language grammar and syntax are preserved and so should be considered home language input.</li><li>• this is extremely hard to answer. re (a) education - yes, up to a point (ISCED levels 0-2) education may be indicative because if parents are barely able to read they won't read to their Children either, but there is no guarantee that very highly educated parents (ISCED levels 7-8) necessarily provide higher quality input to their Children. (c) accent is dicey too - probably some threshold? you can have an accent but extremely verbal, literate, gramm. correct - or not. (d) mixing depends a lot on Community and family Conventions and a lot of mixing is not necessarily a sign of low input quality! (f) digital exposure is tricky too - what sort of language?</li><li>• you need to know if the peers are native speakers</li><li>• These elements may affect input quality, but are not determiners. Input is quality if it is comprehensible</li></ul>

- direct measure of interlocutor language skills is best (if possible)
- They all are, but they also vary in quality themselves, so they may need to have quite broad criteria attached to them to keep them user-friendly.
- As to (c), as long as the child has early access to native speakers and their phonology, accent should not be a problem. As to (d), in highly proficient bilingual societies, mixing is the norm. Whether mixing defines quality thus depends on the community. Children separate the systems even when exposed to mixed input - again, provided they also are exposed to monolingual input.
- I feel like some of these statements can gather valuable information from a linguistic status and information perspective. However, adding the more judgement term of "proficiency" implies that an individual who code-mixes, has a different accent is not "proficient" in their language. Of which I do not agree. my concern is such a statement could guide the administrator in making judgement based observations on something that is completely acceptable (such as language mixing)
- see previous comments
- I think parental education need to be asked but more important would be the communication style which is difficult to measure. Proficiency in each language need to be connected to the language this person uses with the child. Otherwise it is not that important. Language mixing should be asked about but especially with bigger kids not sure about the relevance in terms of quality.
- If a person speaks a language well this does not imply that the person can speak well to young children
- quantifying different speakers and contexts as these translate into different scenarios and vocabulary exposure
- I would NOT call this input quality, as that implies something much different than these variables. That being said, ALL of these are variables that would be useful to include.
- 1) the accent affects the phonological aspect of the input but not the grammatical nor the semantic aspect of it.
- a) education is not necessarily a measure of linguistic proficiency - there are many reasons why a person may not have accessed formal education. c) accent / variety of language is not connected to 'quality of input' (except positively, in terms of providing more additional data.)
- This would require some attention, as exposure to non-native speakers doesn't necessarily compromise language development or lower its quality, especially if the child hears the language from multiple speakers. In fact, one aspect of quality that has to be

included is number of speakers: the higher the number, the better chances for language development

- I have select option 3 for all of the above because these are not yes/no matters - these things MAY or MAY NOT be linked to quality -- plus see note re quality in previous answer
- Again for dialects/pidgins/creoles, there are often no formal measures of proficiency available.
- Whether interlocutor accent is relevant depends on the goal of a study and the dependent variable. If the study is about phonology it would be relevant, but this would be different for a study on syntax. Language mixing is a broad notion. As there are many different ways in which people mix, I am not sure what it tells about quality. Digital exposure is very broad too, it depends much on type of digital exposure.
- (I am assuming here that this is about \*linguistic\* quality...)
- Language-mixing should be considered as a natural part of bilingual exposure. Language-mixing does not mean that the child received "poor" input!
- I wonder whether it's possible to disentangle (f) and (g) from the measure of input quantity, and hence whether a questionnaire should target these areas twice.
- I disagree with mixing because I wasn't sure whether the assumption was that mixed input was poor or strong re quality. I think the child can have weak input quality in monolingual speech and higher input quality in bilingual or multilingual speech. I don't think whether the input itself has multiple languages is necessarily going to tell us whether the input was good quality. It will tell us what the input was like, but not what the quality was like. Similarly, as long as the interlocutor is comprehensible, then I don't see why accent matters.
- I answer "I don't know" when I think that the various points proposed have an impact on the level of language (syntactic complexity for example) but not on the possibility of a good acquisition of the language.
- Language mixing can often be a sign of fluency as it is governed by rules so I do not think it has a negative impact on input quality unless someone is mixing because they do not have a high level of vocabulary and need to borrow lots of words.
- I think we know about this at the present time but not nearly enough to make strong statements. Also research in this to date has focused on a few select languages and cultures,
- This is not exactly clear to me -- for instance, do you mean the higher the language mixing lesser the input quality or vice versa? This needs to be clarified!

- This would give a gross measure of quality with a lot of caveats. For example, we can all agree that parent education is important but does it account for much variance?
- This is exactly why I think a composite score might not be good.
- Peers - depends on the language ability of the peers ...

**Statement 27:** The language proficiency of the child's interlocutors should be documented (based on the respondent's estimate). This should include estimates for:

**27.a: each parent**

**27.b: any siblings**

**27.c: other members of the households**

**27.d: playmates**

**27.e: teachers**

#### **COMMENTS:**

- and updated for c, d, and specially for e
- Extended family (if they have regular contact with them). And one needs to consider the change that can happen over time, so, it would need to be updated regularly..
- I find this hard to answer. Who is going to fill in the questionnaire? How would the parents know the proficiency level of all the other interlocutors? Won't it lead to a lot of guessing? And if you yourself speak a language not so well, it is well known that you may overestimate the language proficiency of others in that language. So I would ask about less here.
- I agree but I think this will be quite challenging and could be a barrier to completion, so such data could be optional and included if available or known.
- (c) could probably be dropped
- I would prefer that any observations/notes on proficiency be done in a non judgement/biased and careful manner. I hesitate to encourage the administrator to ask the guardian questions on what their thoughts were on language proficiency of each of the above.
- In a minority language context such as Irish/Welsh etc the teacher's language proficiency in that language is important as many are L2 speakers of the language
- see previous comments

- As stated before, this is important in concordance with the language used with the child. Unless it's implied that language learning skills of parents in general would be affecting the quality of language exposure (as in education)
- I am not sure how practicalable this would be. I would weight exposure to main caregiver's language as important.
- If parents (anyone filling in the questionnaire) would be asked about proficiency of other child's interlocutors, a clear definition of "proficiency" should be provided.
- As long as each language used with the target child can be included
- Seems unrealistic to do this for every multilingual child
- (b) /(c) and (d) only if the child is in regular frequent contact with these interlocutors.
- I'm not sure how accurate this would be
- but as an overall measure otherwise, it would be so complicated
- incl. other caregivers, as those may be more frequent contacts than parent; should not focus too restrictive on households, as many children spend daytime (communication) outside their homes
- again, if the questionnaire is filled by the parents, it might be hard to judge the proficiency of the teacher as he is supposed to be fully proficient in the language he/ she is teaching.
- These data could be challenging to collect!
- see comments above
- Again, not sure D and E would be feasible...
- again I have chosen 3 because this is not a polar matter . Who will determine proficiency ? . Are you talking about linguistic proficiency? . I regularly interact with highly successful professionals who speak English as a second language . They have lots of grammatical differences in their English but are very proficient communicators
- Difficult to establish the language proficiency of parents/other members of household without frequent contact/use of a test. Could be subconscious bias in response also.
- I am wondering how the proficiency of playmates L1 or L2 could be estimated by parents? Especially in school age children. Especially if parents themselves are L2 speakers. Ditto for teachers. Could parents really be able to estimate a teacher's

proficiency in the majority L2? or would they make binary judgments, like the teacher is a native speaker or not?

- Best to separate between a dichotomous native and non-native - not by a continuum language proficiency.
- I think the last two will be hard to find out based on a parents questionnaire.
- It is relevant but I am not sure how reliable measures are. Estimates depend much on normative ideas/attitudes which can vary strongly across individuals, groups, cultures
- Although I think this is useful, it will be too much to ask separately for all these interlocutors
- First, not all children have 2 parents. Second, it may be more justifiable to have the respondent identify the child's top n sources of input, and consider \*their\* proficiency, whoever they are. Third, it may makes sense to query older siblings but not younger ones.
- Proficiency in which language ?
- I am not sure about the teachers because even though I know from experience this may not be true, surely they should have English language proficiency as a given in order to teach?!!
- I don't know how a respondent could assess points a to d accurately? And do you mean language proficiencies in all languages spoken or understood by the child, or just the majority language?
- And family, friends in other countries that the child as access to
- It would be very difficult to document on the proficiency of playmates and teachers and perhaps create false data.
- (d) May be hard to measure. (e) may be possible to gauge if the type of school is documented.
- See my previous comment about other members of the household. With regard to teachers, I'm not sure all parents would be able to make the call. In fact, for any of these, you will need to think carefully about the ability of hte parent completing the questionnaire to make such estimates. Using functional descriptors made facilitate this.
- Other individuals that the child interacts with such as neighbours, and shopkeepers.
- I think that the proficiency in the L2 is of importance and can be estimated. One can estimate one's proficiency in a new language, But it's much harder to estimate one's

proficiency in one's L1, and I think there are cultural differences in how people approach that

- Should be dependent on the frequency of interaction with that interlocutor
- d) playmates depends on age of the child.
- Playmates are going to be more difficult to track.
- Parents and siblings are more important at younger ages, teachers and friends as they get older.
- Not sure how accurate the respondent will be with these estimates, though. Especially for playmates and teachers.

**Statement 28:** The language proficiency of the child's interlocutors should be estimated in relation to a native speaker of that language.

**COMMENTS:**

- Who would estimate the proficiency in the language which is unfamiliar to the person who is doing the test
- Very important
- not necessarily after long time in an L2 country
- Depending also on the needs and possibilities that are possible for the child/family. What parameters would you choose for English? Australian, US, UK etc. English?
- Just for teachers, that should be certified
- Again, the scale of proficiency should be kept simple and easy to grade.
- Native versus non-native is a controversial topic - what counts native in terms of accent, dialect etc.? Native to that region for example?
- Given that heritage speakers are often native speakers of both their languages, the concept of nativeness becomes suspect. A comparison to monolingual speakers - even if equally suspect - could be used for self-ratings, however.
- This would depend on which language. I would not want to attempt to determine whether a parent/guardian had appropriate language proficiency in a language that I do not speak. again- language proficiency statements outside of DLD/SLP language assessment concern me as they add an element of judgement that can be problematic. Particularly in a healthcare context with differences in cultural/social privilege and power.



- The issue of a native speaker can be hard to measure in some contexts and may not always mean an L1 speaker- the native speaker should be defined
- The term native speaker is an ideal condition. i don't know how far it is comparable and moreover, each language has their own linguistic behaviour
- This might be good but would be challenging to obtain reliably as a required part of a questionnaire
- I don't think being a 'native speaker' is the best indication of expertise.
- Not everyone could do this; special training would be required
- The term "nativeness" is a very problematic notion.
- as a benchmark
- not, if they are bilingual / multilingual
- We should move away from the idea that 'native speakers' are always the ideal models. Instead, proficiency levels based on communicative competence would be preferable.
- given what is known about native language attrition (even due to passive exposure to multilingual environments) and the progressive disappearance of the idealised monolingual native speaker, we should stop using this as a reference point
- We always compare to the native speakers but cant say, in the present days of digital influences on the language. Is it necessary ?
- 3 chosen because this statement makes no sense. Which language are you referring to when you say "that" language ?
- How is the parent, who may have limited proficiency her/himself supposed to be able to judge this?
- Difficult to estimate language proficiency of people who are not well known to respondent. Many native speakers aren't fully proficient in their own language.
- I guess this question is going to make us start a wonderful debate
- A native monolingual of that language? Not necessarily. It should be estimated, as much as is possible, on a "typical" member of the population representative of the person completing the questionnaire.
- I don't think we should use the term native-speaker - what do we mean by native-speaker? Who is more of a native speaker of French - a parisian or a quebecois? Slippery ground here ...

- For many minority languages, native speakers are bilingual. Do we therefore compare proficiency to the 'standard' language even if it is not spoken by the majority? I think it should be estimated in relation to a native speaker of that language IN THE SAME REGION, understanding that the native speaker also speaks other languages/dialects.
- as well as to other children in the community
- This is likely the easiest and most consistent way of doing this.
- No - I think it should be estimated in terms of comprehensibility, not whether the interlocutor sounds like a native speaker
- Depends for what the reason for such estimation is. I think it is enough to stay with broad qualifications, such as low proficiency, medium or high, We should not be getting into arguments about near nativeness and such
- We have native speakers who have not spoken their language in ages so we will have to be really careful on how we do this. As it might mislead us
- This measure of language proficiency requires discussing the concept of 'native speaker'. There are some languages where there are no monolingual speakers for the minority language. Who represents the native speaker in these cases? 'Heritage language' is seen as incomplete if it is evaluated under the concept of 'native speaker'. Furthermore, the dialect of the language must be considered.
- I am not sure native and near native are the relevant measures. I think a simple question on how proficient one is, who yield more accurate results as even highly proficient bilingual speakers might not see themselves as near native, but this distinction is less relevant for the input the children get.
- Hmm. I see the problems with have a "native speaker" standard, but I'm not sure what other realistic options there are, especially ones that can be applied for a variety of different contexts and languages.

**Statement 30:** Language mixing (in interactions with the child) should be documented for each interlocutor.

**COMMENTS:**

- With guidance that code switching should not be discouraged
- depends on the mixing...: frozen phrases, simple lexicon, or grammatical structures (subject omission...)
- This can vary from situation to situation, and depends on the phase of language development

- I wonder how well people know this about themselves and other interlocutors
- Too difficult to measure objectively
- In an ideal world, but the more I see the proposed complexity of this survey, the more I'd argue for these aspects to be desirable and optional rather than required otherwise they become cumbersome and a barrier and fewer practitioners would make use of the resource.
- Much research has shown that mixed input is irrelevant.
- Again- this should be gathered for clinical information purposes, and clearly stated that it is not for language "proficiency" purposes.
- It might be excessive for each interlocutor- certainly parental/ peer input and child use should be measured. Would have to weigh up against the other questions
- I think language mixing is quite normal when more than one language is spoken. I think there needs to be some balance between ever more detail collected and the user-friendliness of the questionnaire
- Not sure this is always relevant.
- Need to be thought through in which accuracy can it be documented (e.g intra-sentential, inter-sentential, dominance, frequency)
- Of little importance in my view
- Not for each interlocutor
- yes, language mixing differs according to the listener
- needs to be measured (e.g. using video), since self-assessment rather invalid, language mixing less noticeable
- The question is: how?
- Not something to worry about.
- however what significance are you going to attribute to language mixing ?
- But how? A predefined measure needs to be included.
- Only for those that would mix languages (e.g. parents, siblings, friends).
- Difficult to document in terms of time and funding.

- At least for the main interlocutors at home and the closest friends/family members outside home.
- Not sure how easy this would be to document in a questionnaire or what purpose it would serve. Especially of mixing is going on for different reasons. Mixing by two parents each fluent in both languages would be different from parents mixing when speaking their L2 because they have lexical or grammatical gaps.
- It would be good to include some measure of mixing to be able to document whether it is part of the input, although I am not sure what it tells.
- Probably only in studies that investigate code switching
- I don't see why language mixing would fundamentally alter any patterns that might already be predicted by the reports of language use and exposure. This doesn't add additional useful information.
- For the same reasons previously noted about measures of output
- May not be easily quantifiable but parents could answer whether or not they use multiple languages with child at once and factor that into their overall percentages of % exposure to each language.
- And by language, it is important to include dialects.
- If "documented" refers to self-report, then this measure is likely to be very prone to error and different types of error depending on both cultural and individual differences.
- only for the main interlocutors
- To the extent that parents are a) aware of this and b) able to make a reliable estimate. Again, this is perhaps an empirical question which the questionnaire you're developing might facilitate the answer to.
- It could give you an idea of whether the child mixes languages and whether they are usually in bilingual or monolingual mode with each interlocutor.
- I think this may be too difficult for respondents to document. Bilinguals are often not aware of their code switching
- How would that happen though? I think there are tools already that can do that.
- It is interesting for sure but what does it measure.

- Depends how extensive the mixing is. If it's occasional insertional mixing, it's probably not that relevant. But if it's constant mixing throughout the sentence or interaction, it might be more relevant.
- language mixing can be very difficult to interpret as it is very common even when language use is good

**Statement 31.1:** Language mixing should be estimated (in terms of exposure and use).

**COMMENTS:**

- (See previous comment)
- Same comment as before.
- It depends on how this is framed. Several communities only have a mixed/code switched variety and yet see this as imperfect.
- I don't quite understand this question. Do you mean that the child's exposure to mixing and the child's use of mixing should be estimated? Or whose mixing?
- Not ideal, but may be the only practical way - if this is the main subject of a child's assessment, it may need more extensive exploration.
- this could be done in order to gather more data on the role of mixing in bilingual development - as to input/output quality, it is not very relevant
- There are a lot of questions here on language mixing. I do feel like this is important information for clinical purposes, but it brings some serious questions when linked to the concept of linguistic "proficiency"
- yes as per previous answer
- across all interlocutors? It is not clear what this question means.
- this is not always bad. Some communities do this regularly and routinely
- I think this will be difficult for a parent to be able to do this
- not sure how to interpret this question
- unclear: whose language mixing? Exposers or child, or all?
- It depends in how much detail; broad categories (never, sometimes, often) would be sufficient

- In bilingual contexts, language mixing is a very natural phenomenon. therefore, it is important to check whether it occurs or not (if the parents think they code-switch or not) and how often (frequently or not frequently or not at all). any other measure might be hard to gather.
- see question above
- If possible
- Very difficult to measure. Almost all bilinguals mix languages.
- should be observed
- I suspect much more language mixing occurs in the home setting therefore it would be difficult to estimate with any degree of accuracy.
- In broad categories.
- I ask parents which language is dominant when they mix and I ask if they can estimate in percentage how much they use each language when they mix. That is mostly doable for parents.
- Is the question about quantification?
- Not sure how you could measure this accurately and how much more it would tell you.
- Language mixing is natural to multilingual environments and is often a positive thing in communication. Fluent mixing requires good language skills.
- Yes, and which languages are being mixed and in which contexts
- And compared to the 'norm' of that particular region.
- I don't know what this would show?
- and what type of language mixing
- This is a great skill and one that highlights higher cognitive ability and can demonstrate the speakers ease of living in two languages and cultures.
- This might be a cumbersome task.
- Very interesting if you are interested in codeswitching but of less relevance as predictors of the child use, unless we are interested in the child's language mixing.

**Statement 31.2:** This should be done through examples (of the different types of mixing) rather than overall estimates from the respondent.

**COMMENTS:**

- I did not understand this statement
- Definitely
- Code switching is something that a lot of people seem unaware / report that they do not do but can be observed ++ within sessions
- (See former comment)
- Self-reporting is going to highly unreliable in this case.
- It would be great to implement this! One potential difficulty is creating comparable examples of types of mixing across the 13 languages.
- I strongly feel that asking parents/guardians questions with regards to code-mixing/language-mixing is inappropriate and would likely risk an element of communicating judgement and bias for something that should be considered natural and a normal aspect of bilingual/multilingual linguistic life.
- some examples can be useful, but perhaps overall estimates could be included also
- i am not quite sure, if a speaker would explicitly provide examples as when there is a conversation, we normally subconsciously adapt to language mixing.
- If language mixing is going to be estimated (which I'm not sure about), then it should be done through examples indeed.
- I would probably rather have something that can be calculated but examples could be extra to give a qualitative view.
- it could be both
- types of mixing should be also predefined (with provided examples).
- same comment as above
- relevance then unclear; if relevant, then consider all, otherwise: biased
- Not sure if this is necessary or possible to accurately report
- If possible.

- difficult to provide a comprehensive inventory of types of mixing which are used in different bilingual settings, especially considering the difference between bilingual communities when mixing is common and monolingual or diglossic communities where mixing is frowned upon
- It nice. This question made me to think of what examples can be cited?
- should be observed and objective
- These examples would have to be region specific.
- I really wonder how well parents could respond to questions like this.
- A separate language tool could measure it through examples. An overall estimate should be sufficient.
- I think it can give good insights, but I also think it is hard for most parents I see to give examples
- too detailed
- The template questionnaire should include examples/explanations of the following: alternations, insertions and congruent lexicalizations (for definitions and examples of each, see 'Bilingual Speech: a typology of code-mixing' (Muysken, 2000). In these examples, you should be consistent with your unit of measure (e.g. always a clause or always a sentence - I would go with a clause). When you translate questionnaire in 13 languages, it would be good (although hard, maybe impossible) to provide an example for each of language mixing types and potentially from existing corpora for those languages. Otherwise, clearly define what you mean by each of the above three types and point to relevant literature so that whoever uses the questionnaire and with whichever language combination can come up with adequate examples before distributing it.
- overall estimates might still be useful
- I would think both as they provide different kinds of information.
- If possible, but again, this depends greatly on the linguistic awareness of the respondent...
- With the same caveat as above, namely that it's not clear to what extent parents will be aware of this and/or be able to reliably make the estimates.
- I think it's way too hard for most people to report on this
- If we decide that this language mixing is of interest, then this is the way to do it.



- |   |
|---|
| <ul style="list-style-type: none"><li>• Too hard to maintain any sort of consistency otherwise. The respondents will probably have very different ideas of what mixing is. Or not consider such situations mixing at all.</li></ul> |
|---|

**Statement 32:** Language mixing should be documented separately for language use and language exposure.

<b>COMMENTS:</b>
<ul style="list-style-type: none"><li>• If possible</li><li>• Again I don't quite understand the question. Are we talking about the child's use of mixing vs the child's exposure to mixing? In what way is the question different from Qu 31.1?</li><li>• It's more of a subset of exposure and use, so may complement those better if included but highlighted under these banners.</li><li>• In the child - see above (was this the same question I just answered?)</li><li>• Probably</li><li>• Using more than one language is part of the process of learning an additional language</li><li>• Again, to the extent that language mixing is going to be estimated, this should be done separately for use and exposure.</li><li>• This should also be measured over time and with each person.</li><li>• The question is not clear! Language mixing by parents? Language mixing by child?</li><li>• I think language mixing is part of the natural use/exposure</li><li>• you mean documenting by the interlocutor if they mix or not?</li><li>• see comment above</li><li>• possibly - assuming one knows exactly what the different estimates would provide</li><li>• Paradoxical.</li><li>• BUT what significance are you going to attribute to the findings?</li><li>• Difficult to establish?</li><li>• Not worth it</li></ul>

- Also, the second round should include a statement about whether language mixing should be documented separately FROM language use and exposure or be considered as a part of it.
- I just am not seeing what useful information this provides over and above the other use and exposure variables.
- Language mixing in the input is more relevant, although even here I would be surprised if the final estimate would differ between instruments that did vs. did not ask about language mixing.
- Possibly. It depends how this data might impact on overall findings. I am not sure how easy it is to separate them out and if you do so, how accurate that would be. However it would be useful to know if a child hears and understands but never produces a certain language.
- Do you mean exposure to language mixing?
- Again, I'm sceptical about how credible this will be.
- Same proviso as previous questions on language mixing
- This is a bit tricky.. use I mean. exposure yes.. use not so sure. Unless this questionnaire is aiming to estimate productive skills as well as exposure.
- It can be quite different.
- I think that the child's tendency to code switch is easy to document and is telling, so should be documented
- I don't know how realistic this is in terms of quantity. I think determining whether it happens at all is going to be hard enough.

**Statement 36.1:** The literacy practices of the parents need to be documented (e.g. print exposure, reading the news, reading novels, etc.).

**COMMENTS:**

- This is likely to link to parental education levels but does it consider the input the child is directly receiving?
- As a cultural level, but more important it is the quality and time of contact with children
- Depends if the community is literate. There is also the issue that religious texts may be 'read' in another language, with no understanding (or productive use), such as Pakistani heritage UK speakers using Mirpuri, Punjabi or Urdu as the home language, but able to recite the Quoran in Arabic (by rote) despite having no use or understanding of Arabic.

- this can offend however
- I would find it more important to learn about the parents' literacy practices when communicating with the children (e.g., reading to/with children) than the parents' practices on their own.
- I think this can be captured by parental education.
- perhaps the effect of this is already sufficiently taken into account when examining the quality of the exposure
- Even though this would give information about education, it might also be an embarrassing question and it might be highly inappropriate in a mass shelter situation.
- There is some controversy in research on the accuracy of literacy/education on vocabulary/language development of children. I'm not sure the literacy levels would add enough value to risk the potential distress/damage to clinical relationship asking such questions would cause
- It is more important to measure in the child so for the sake of reducing the amount of questions I don't think it is the most important
- if the questionnaire is to help diagnose a language disorder then I do not think parent's literacy is relevant
- It will correlate with the education level
- Since this is very much in connection to the time parents actually have for those things and not necessarily to how much they would want to do that, not so important. Attitude towards literature matters of course and the opportunities parents offer to their children in this sense.
- This would need to be sensitively worded/considered but knowledge of carers' literacy in L1 is helpful to know.
- I believe it is more important to document whether parents transmit literacy /whether parents read to the child.
- As an SLT it is useful to know if the parents are not literate but I think the previous questions regarding parental levels of education may cover this. I don't think I need to know this level of detail
- and differentiate "the parents" and other caregivers
- This is a separate issue

- If by literacy practices you mean the books/ documents themselves, then with all the new technology, a lot of parents do read however, on the screens.
- Again, prioritising the length of this measure and practicality of implementation.
- I have chosen 2 because I do not know your intention in asking this , some languages have no written form - will this then discriminate against those families ?
- 
- Language proficiency of parents is enough to screen for language "quality" of parents.
- I'm not sure how different this would be from parent education in terms of its predictive value. I also think that it might embarrass parents with low education
- I think it is interesting to know, but when the questionnaire is too long, we don't have enough time for it. So then I would let this one out.
- too much information for a generic questionnaire
- I would only want to capture practices that are relevant to the child (e.g. bedtime stories in L1 vs L2)
- Ask the parents about home literacy activities conducted with the child seem sufficient to me. I don't think there is a need to ask about their own practices.
- I strongly agree but I am not sure how easy this will be to gather this information without sounding judgemental.
- Also reading religious texts. Do you mean print literacy here or orality as well?
- For what purpose? To infer quality of input?
- Unless it's input directed to the child, a proxy for this could be parental education or proficiency in each language. It may not be necessary to have this level of granularity here.
- Unclear to me whether the benefit of this outweighs the risk of appearing to evaluate and judge their literacy practices.
- I think this might be intimidating and potentially cause offense for some parents.
- In Western societies at least, education level gets at this. There are cultural differences in what people read. I see this as culturally biased and having the potential of coming across as judgemental or arrogant

- This question may be a sensitive issue in communities without literacy. Perhaps there should be a guide on how to ask this question in these cases. Questions about the educational level can also be a sensitive issue in these communities.
- This might be a better proxy for input quality than the measures suggested previously.
- Why are we interested in this? If it is a measure of education, we can ask directly about education. If it is about the model the child sees at home, it is a different study. If it comes to measure proficiency in the different languages, a direct question is better.
- could be interesting but is very dependent on the different cultures, so it is difficult to interpret

**Statement 36.2:** This needs to be done independently of parental education and socioeconomic status.

**COMMENTS:**

- We know there is a correlation between education and literacy practice, we don't need to reinvent the wheel
- (Previous comment)
- see above
- see above answer. I don't feel any benefit outweighs the potential risks in clinical relationship/distress to parent
- SES needs to be measured anyway
- as above
- If the question preceding this would be asked, then yes.
- as long as storytelling and reading are considered equal activities, same for picture books
- I think fine to have a measure of SES, but should be separate.
- poor question in first instance - see question above
- I am not sure because these are all such indirect measures
- May be interesting way to get at protective factors - parents with low SES/education who read more tend to have children with better outcomes (hypothesis)

- Are you going to judge SES in home country separately from present country of residence? Very important in terms of refugee families where parents from all SES backgrounds may have missed out on education.
- I understand that literacy practices and SES may not always correlate, but the question is whether this level of granularity is necessary. Perhaps it's culture specific.
- I disagree with the premise of the statement (because I don't think this should be measured at all).
- If we are making assumptions - we can assume higher parental ed leads to higher SES and would mean more print exposure in home. But presumably it's best not to assume.
- I think parental education and SES can give you a good indication of the parents reading abilities.
- I don't think this should be done
- Yes, to be able to compare different cases. However, there should be a guide on how to ask these kinds of questions in communities where these issues can be sensitive.
- What are we measuring by it?
- Unless there is clear research for numerous societies showing that SES and literacy practices correlate highly.

**Statement 36.3:** Parental print exposure is an informative measure of literacy practices.

**COMMENTS:**

- Will they printed in their own languages
- Type of exposure may be but not necessarily amount
- (Previous comment)
- Some studies show that just having more books in the home = better literacy for the child
- Even if this is the case, there are problems, see above.
- as above
- Probably depends a lot on child's age. With smaller children more important what is being done with the child, maybe parental print exposure more important with school aged children.

- see above. Some languages do not have written form . you risk passing judgements on literacy when in fact parents have not had need of same
- I am absolutely no expert in this, but I think this used to the case (such that number of books in the home also was a good measure) but I'm wondering whethr this is still the case nowadays, now that people use e-readers, and phones and tablets to read books and newspapers...
- It usually is but there may be a lack of printed material in the parent's home language which affects this response.
- There are arguments that orality needs to be recognised as well??
- A bit unclear; literacy practices for parents or for children?
- this depends on what you want to use the question for
- It is definitely one measure of literacy practices but not the only one
- I am really not sure about this part. The link between Parental print exposure and literacy practices might not be that clear/evident
- I think this is an empirical question.
- There is a strong culture of oral literacy practices such as oral storytelling and this should not be overlooked.
- Parents may read a lot to their children even if they do not read their own books etc.
- literacy practices do not apply to all cultures and languages
- Not necessarily. There are good readers with a low formal educational level. There are cases of people who drop out of school because they need to work, not for lack of passion for knowing and learning.
- I'm not sure I understood the question. I think it is ONE of the measures of literacy practices
- It is a measure of their literacy practice.
- Not sure what is meant by print exposure different from reading novels, reading the news, etc.

**Statement 38:** If a parent was educated in more than one language, this should be documented separately for each language.

**COMMENTS:**

- They could be used as an interpreter
- If we measure proficiency and use, we have it captured
- Alongside the interactions the parent has with the child in which language - a number of my families have parents educated in a range of languages but they do not use this range with the child
- Yes. The independence of languages or the degree of mixing or "semilingual" competence is relevant
- Thank you for including this! Parents who speak/read/write multiple languages often have studied and worked in multiple languages too.
- this could be relevant
- I don't see why - the parent might be educated in languages that are never spoken around the child. If you are referring to education in languages spoken around the child, the question is more relevant and I would give a higher score.
- again- for my clinical purposes this would not be a question that would provide enough value for language profile in order to risk the clinical relationship with the client/parent/guardian
- It has been suggested that the language of education affects especially that language in child's development.
- Also: country of obtaining the education may be important (especially for new migrant parents)
- I think the level of proficiency of parents, including reading and writing skills can be enough.
- should not only address "education" but raised in diverse language settings, grown up with more than one language, ... please remember that the native language of a caregiver may not be the best (in terms of quality)
- This would be useful
- both raised and educated
- Again, I think this is more relevant for a demographics form
- Are you asking about literacy here ?



- Perhaps this question could be skipped and the info obtained through proficiency ratings?
- Yes, this data is important as it creates multilingual parents and their practices, vis a vis, monolingual, should be documented
- Maybe ask about parents' language proficiency in each language spoken
- See work by Erika Hoff and others showing the importance of doing so.
- by document do you mean what kind of educational experience they have in each language? Or 50% of the school day in Language A and 50% in Language B? both would be good to know.
- There is some research evidence that this does indeed matter
- The evidence for this as a predictor of the child's language is limited.
- What is this a measure of? Proficiency? Ask directly about speaking. reading and writing.
- Depends a bit what the measure is used for. If it's for education overall, then maybe not relevant. But if it's related to language proficiency, then it is relevant.

**Statement 41.1:** There should be a question about the status of each of the child's languages

**41.1.a: within the local community (including school)**

**41.1.b: in the "home" country (if applicable)**

**COMMENTS:**

- No language should be afforded a higher status by professionals as they are valuable for different reasons for the child - attitudes towards the language by the different groups should cover this
- in particular for negative opinions and prejudices
- perceived status?
- Again, this is a different question
- Not sure this is achievable with any consistency or reliability.
- This may be covered by the attitudes question

- No sure what "status" refers to
- I am not sure what 'status' refers to here.
- Referring to the answer before, these might be tricky questions to answer objectively.
- "status" should be defined for persons filling in the questionnaire
- Not sure how your "attitudes" Q differs from your "status" Q
- if people know and care enough
- (b) home country is only relevant if there are contacts with the home country.
- I don't know what you mean by status.
- incl. former/past languages of use
- 'status' can mean different things to different people. The term should be spelled out
- For dialects/pidgins/creoles, the status is often low in the general population but it is important to know when parents agree with or push back against those sentiments.
- Will parents necessarily know this?
- Should be a continuum
- Also so you know whether there is any form of written language.
- I am not sure what is meant with "status" and how this would be asked in a comprehensible way.
- I like thee 'if applicable'. You can add this to many of the earlier statements.
- I would separate out school and community and home, and keep home country as separate still. E.g. home may not value heritage language and be keen to only speak majority language at home.
- See previous comment.
- Status matters in terms of development
- I think this is info that is available to the user of the tool
- And also within the family

- I don't think parents are necessarily best placed to make this kind of sociolinguistic judgements.
- Do you mean the respondent's perception of the status? Or some independent measure of the status, for example as determined by literature from sociology?

**Statement 42:** There should be a question on attitudes to language mixing

**42.a: within the family (at home)**

**42.b: within the local community (including school)**

**42.c: within the broader society**

**COMMENTS:**

- Not sure what questions about attitudes really add. In my own research they never correlate with anything.
- With guidance that this is not something the child should be discouraged from doing!
- Important and very personal. Manifestation of the social status, situation, values and aspirations
- And he extended family...
- Attitudes are a whole different question
- My previous comments apply to b and c here due to inconsistency and reliability factors
- Attitudes to code-mixing might be revealing though they only need to be evaluated if mixing is kept as part of the quality of exposure/use measures
- This question would be more to guide counselling with parents on encouraging them not to feel ashamed or to for them to discontinue language mixing.
- it's a useful discussion to have with parents in an advisory way
- Not sure what is meant by language mixing
- but it implies it is not a good thing which is not always the case
- This is interesting from the research perspective. Yet, when capturing child's input and exposure, the frequency / presence of code-mixing is more informative, than attitudes to it.
- such a questions may connote language mixing to be perceived as a stigma

- Separate questionnaire
- again, making clear what 'attitudes' means
- Not sure if it'll add any useful info
- this suggests mixing is option and sociolinguistic - neither of which are true
- Again interesting, but why exactly would this be relevant?
- If it is relevant. In one-parent one -language families, community and society are not relevant
- Very interesting!
- I disagree with (b) and (c) for the reasons given in response to previous questions on attitudes. With regards to (a), I don't think this will be relevant to all and hence if it were included, I would only have it as an optional question.
- See my prior comment on attitudes... it depends on how many assumptions we're prepared to make.
- In case of multilingual societies it will not be a relevant question.
- All these attitude questions are great for a supplement on attitudes which can help interpret the results of the exposure/experience/use part. But I miss the child's voice on all of them
- Not sure how useful this is. But maybe because most of my experience is in not-much-mixing contexts.

**Statement 43.1:** Parents should be asked if they feel pressurised to speak the societal/majority language.

**COMMENTS:**

- I have had comments within my school, that all EAL children and their families should learn English before they arrive
- Important to ask about speaking the standard variety of the language: often it is socially/politically not acceptable to use the standard dialect/variety while it is the standard variety which is/should be used in school.
- Better to know their attitude as a reaction to it (previous question)
- this might be useful to find out

- It is doubtful that this question will be answered frankly if the parents indeed feel pressurised. It might be better to ask outright if they have been advised to speak the majority language at home.
- I'm unsure if I would word it that way. But in my experience there are differences in views on language depending on whether the family is from a minority or majority context (ie someone from Québec will advocate strongly for as much services in French and highly value French in an anglophone province, while someone from France may simply want their child to speak English)
- This gives an indication of any other advice the family may have received
- if they don't know obviously they won't speak. If they can and is the need of the hour they will surely converse
- Yes, many parents are "advised" by school teachers / SLPs to switch to the societal/majority language.
- This should be a separate questionnaire
- This will not quantify the bilingual abilities of the child, but I feel it is more related to the status of bilingualism worldwide.
- depends on where they work and social class.
- the choice of 'neutral' questions is essential here
- Depending on the research aim, this could be interesting to know. Perhaps, in cases where there is a risk of socially desirable answers (parents reporting they speak the majority language, while actually they don't) this would be worthwhile
- This would only serve as an indirect indicator of attitudes towards languages used, wouldn't it?
- Several parents I work with have been told by daycare educators that they need to practice French with the child to prepare them for French school even when parents are not very proficient in French
- I don't think this is necessary as part of an input quality/quantity questionnaire but may be necessary as part of an attitudinal questionnaire.
- Optional question ok, but not as obligatory question.
- I'm not sure a question like this is an index of exposure. It would be useful if someone was researching attitudes etc. but if it is strictly exposure then not sure if this matters so much.

- Many parents are told they should do this and it may be detrimental to their child's acquisition of both the minority and majority language.
- I think you have to specify pressurized to speak it where. Immigrants in general view learning the societal language as a necessity, which it arguably is. It's another question whether they feel pressurized to abandon their L1 in the home
- And the reasons thereof.
- But also if they or the child feel pressured not to speak another language in any context
- It is an indirect measure of the previous questions. If we have a supplement on attitudes, it is a great question
- It is a delicate kind of question to ask

**Statement 43.2:** To identify the source of that pressure (if any), a list should be provided for the parent to choose from.

#### COMMENTS:

- with options to add items not listed
- Not sure how that will be quantified.
- Useful info to have when developing locality training
- Delicate question... They will pleasure the interlocutor in the answer...
- with room for 'other'
- maybe - or use an open question here?
- Could be controversial for parents to specify - may be reluctant to answer. Wording may be important here to avoid 'blame'
- Again, many parents do not fully understand the anonymity of the survey. Such questions might induce even more pressure or fears.
- I don't think this would add enough value
- It could be open ended?

- I'm not sure there should be questions about parents feeling pressured to speak the majority language, but to the extent that there are such questions, then it should be through a list.
- It should also include parents own beliefs and feelings as an option.
- but it could just be discussed
- Not sure to what extent it is informative for quantity / quality indices. But it is an interesting research question.
- Although we need to be careful to avoid leading questions.
- a list what exactly?
- no lists please - who has the right to compile a list ? freedom in report needed
- Yes, list + other-option (as list might not be exhaustive)
- 'Other' option should also be offered in case parents wish to identify a different source
- and maybe some other comment options to be listed by themselves
- I am not sure what kind of list that would be.
- including own pressure!
- A list of sources?
- If such question is asked, then a list makes sense with the additional option of "Other" so that they can add their response.
- Maybe an open question is better?
- I only agree in the sense that comparability is good and such lists also facilitate user-friendliness, but I don't in fact think that this question is necessary.
- This can be an uncomfortable question and parents should come up with their unique responses.
- It would be really useful to identify this so that measures could be put in place (such as training of teachers of health professionals) to stop this.
- I would leave it open ended

- However, there are also internal pressure of conforming to the society's model of success (speaking the majority language) and being aspirational. Such internal pressures are common among immigrant bilinguals and it needs to be documented too.
- If we want to know something, a list is good. The question is whether this is important. I think it is not.

**Statement 49.1:** The questionnaire should be available in different lengths.

**COMMENTS:**

- Why?
- I feel that some questions will be impossible to quantify and will be used for clinical purposes, whereas a core part will be used to derive a % of exposure, so yes
- If it is content and language clear, not necessary (good to do pretest trials...)
- Should have triggers for different children / professions. E.g. Children with suspected language disorder should have a more in-depth assessment. The authors should liaise with professional bodies to ensure that professional guidelines reflect this use, or services / professionals may feel under pressure to only do the shortest version by managers eager to economise.
- perhaps two lengths - full and abbreviated
- How will items for shorter version be chosen?
- The most predictive questions should constitute the core. Then additional questions can be added from a larger version in order to explore specific areas.
- as long as it doesn't sacrifice the calculation/information. Although I could envision determining the core questions to create a more concise questionnaire.
- Should be the same for all participants in a particular study. But a standardised questionnaire with different modules would be useful for researchers.
- Short forms can be useful for screening, or to send to families in advance of any face to face assessments to complete in their own time. Then more probing questions can be asked if necessary.
- it might be useful to have quick version and a fuller version
- This is a great idea but most respondents would probably opt for the shorter version
- It's a good idea, but I have no strong opinions on this



- For clinicians could be offered a shorter version that could still be used with the calculator. Otherwise it's useless.
- Absolutely, but deciding how to make it shorter will be a difficult task :)
- how would choices as to what to leave out be made
- Depending on the goal
- Yes but should be made very clear that they serve different purposes and not same data/interpretation possible. Shorter versions unlikely to serve same purpose/aim as longer versions.
- depending on the choice of questions
- Either the questions are needed, or not. If they are needed, then they should not be omitted. If they are not needed, they should not be on the questionnaire. Any respondents who require support in answering the questionnaire should be able to get it with online or face-to-face.
- depending on the bilingual setting/context?
- I think this would be a good idea. For a battery of questionnaires for parents, a shorter questionnaire may sometimes be the only reliable option. For others an interview/longer questionnaire would make more sense.
- you either need the information or you don't. Why ask questions that could be dropped in a shorter version?
- Yes, see my points above (perhaps the parts on developmental disorders, mixing, and attitudes could be made optional)
- It could be helpful to have a quick screener + then a longer version for more in-depth evaluation.
- Coming up with the shortest, yet accurate and detailed, form possible and using it under all circumstances serves its reliability better, I believe.
- How would you get comparable data if questionnaires were available in different lengths?
- There could be a core section and then modules that can be included or not included. Otherwise, the questionnaire might be so long that it will be a deterrent to parent participation.
- I really like that option.

- With flexible components, yes.
- Yes, in reviewing a lot of these questions, I keep thinking - well it depends what the purpose of use is. Perhaps a longer version for certain types of research and shorter version for use in clinical settings.
- I think I would rather see a uniform set of "modules", each with a different purpose. Researchers could select the modules that are important for their questions, but within those modules the instrument would be consistent.
- Depending on the purpose it may not be relevant to include all items and it would be helpful to make a selection for a short version
- But then you won't the same data for each child and does than not depend on the child's situation. The questionnaire should be able to skip to relevant sections depending on the experience of the language users
- Perhaps a screener to determine if there is a need for the long form? Maybe there is very little exposure to other languages and in that case, it would not be necessary. However, if there is sufficient exposure, then I think that there should only be one length.
- Which parts can be removed without impacting the validity?
- I think this is an empirical question rather than a preference. If we are aiming for the highest level of comparability across studies that end up using this questionnaire, then I think it would only make sense to have one version.
- please make one questionnaire with all the questions that are relevant and necessary.
- Perhaps a core part with further add-ons? Whilst I can appreciate that some questionnaire-users (e.g., SLTs) will want a shorter version, I think it's important to maintain some level of consistency across users. Otherwise, it's kind of defeating the purpose of the whole enterprise.
- Not sure what this means. Do you mean a shorter version with a subset of the questions?
- Yes but why? I think it might be useful if there were core questions that was a baseline - and then additional questions people could use depending on the context/purpose of collecting the data were.
- Care has to be taken to ensure that no pertinent questions are missed in trying to keep it short.
- Some parts may not be necessary for all e.g. if a child has only lived in one country.
- Yes, this is crucial for different experimental designs, researchers' needs, etc.

- that would necessitate some extra validity studies, but it might be handy. I think no version of it should be extra long
- Maybe, depending on how people react to activities. I prefer working on something focus on it and finish not do it in pieces
- Yes, as long as the length of the questionnaire does not prevent comparison between communities.
- I think that the existence of a reduced version, similar to a screening, is desirable, although its scope is less
- This is definitely true if we envision different users.
- Not sure what this would mean - how can all the questions be included if there are different lengths. But this raises the issue that, while all the things discussed here would be ideal, a questionnaire covering all this in the detail implied here would be difficult to get people to complete. Also, it's unclear how much of that data would be used for analysis in any given study. But still it would be good to have it available.

## **Supplement 5**

**Paper title:** How to Quantify Bilingual Experience? Findings from a Delphi Consensus Survey

**Content:** Instructions and clarifications for completing round 2 of the online survey.

## **Q-BEx Delphi Consensus Survey Round 2 Clarifications and Instructions**

Dear panellist,

Here we outline some clarifications and instructions for the second round of the online survey.

### **Clarifications:**

Round 1 contained 53 statements, many of which included substatements. Agreement was counted as a sum of 'agree' and 'strongly agree' responses. In order to reach consensus, a statement needed to reach an agreement of at least 75% overall.

Round 2 contains only those statements (or substatements) that did not reach consensus level but may have the potential to do so. To be included in Round 2, a statement needed to reach at least 60% (but less than 75%) agreement. We also included statements that reached at least 60% overall agreement in a special subset of respondents: those who identified as both a researcher and a practitioner.

Round 2 will invite you to re-rate these 'grey area' statements in light of everyone else's responses. In many cases, you will also be asked to rate a reformulation of the original statement.

The report which was sent to you by Bissera contains the distribution of responses only for the (sub)statements which need to be re-rated. After the second round of the survey is completed, you will be emailed a detailed response distribution for all the statements from both round 1 and round 2 for your own record. The detailed reports will also be included as online supplements in the final publication.

### **Round 2 Instructions:**

Following this email, you will receive an automated email by 'Jisc online surveys' containing a link to the round 2 of the Delphi survey.

The general instructions for Round 2 are the same as in Round 1: You will be asked to score each (sub)statement on a 5-point scale, indicating your level of agreement with it. "I don't know" can be selected to express lack of an opinion or lack of specific knowledge. Under each statement, there will be a box in which you can add comments if necessary. Please do not identify yourself in the comment box.

There are three main differences in this round:

1. You will be asked to re-rate the (sub)statements *in light of everyone else's responses* (which may or may not sway your mind). You will therefore need to refer to the report from Round 1 (which you received from Bissera) to see the original responses.
2. Some of the (sub)statements will have two versions: (a) a version identical to the one in round 1, and (b) a reformulated version based on the comments made in round 1. In such cases, you are asked to rate both formulations. For an example, see Figure 1 below.

## Page 15: Statement 32

This part of the survey uses a table of questions, [view as separate questions instead?](#)

<p>Original round 1 statement Needs to be re-rated</p> <p>Statement 32: Language mixing should be documented separately for language use and language exposure.</p>	* Required				
	1 = strongly disagree	2 = disagree	3 = I don't know	4 = agree	5 = strongly agree
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

This part of the survey uses a table of questions, [view as separate questions instead?](#)

<p>Round 1 statement reformulation Needs to be rated</p> <p>Statement 32 [reformulation]: Language mixing that the child is exposed to should be documented separately from the language mixing that the child produces.</p>	* Required				
	1 = strongly disagree	2 = disagree	3 = I don't know	4 = agree	5 = strongly agree
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Figure 1.** An example of the original round 1 statement and its reformulation

- Round 2 also contains some new (sub)statements, which emerged from the analysis of round 1 responses. Please rate those as per the general instructions above. Furthermore, sometimes we provide you with the surrounding (sub)statements to give you a relevant context. Agreement rates for these (sub)statements are provided in square brackets and they do not need to be re-rated. For an example, see Figure 2 below.

## Page 2: Statement 4

**Statement 4.1:** The questionnaire should contain a section to identify children who might be at risk for a Developmental Language Disorder. [Agreement: 77%]

**Statement 4.2:** This should include:

4.2.a: early language milestones [Agreement: 84%]

4.2.b: hearing difficulties [Agreement: 80%]

This part of the survey uses a table of questions, [view as separate questions instead?](#)

<p>Original round 1 (sub)statements Need to be re-rated</p> <p>4.2.c: consistency between siblings' language development</p> <p>4.2.d: issues related to trauma, attachment, prematurity</p> <p>4.2.e [new substatement]: family history of learning difficulties or speech/language delays</p>	* Required				
	1 = strongly disagree	2 = disagree	3 = I don't know	4 = agree	5 = strongly agree
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Surrounding (sub)statements  
Do not need to be rated

New (sub)statement  
Needs to be rated

**Figure 2.** An example of the original round 1 (sub)statements, new (sub)statements, and surrounding (sub)statements

**The deadline for completing the survey is Tuesday, 7 July 2020 (23:59 BST).**

The whole survey should not take more than 1 hour to complete. Note that this round is approximately half the length of round 1. You can complete it in more than one sitting if you choose to do so. In case you have any questions for us, please email us at [qbex@leeds.ac.uk](mailto:qbex@leeds.ac.uk)

Round 2 respondents will be listed as co-authors on the final publication (unless you wish to decline).

Thank you so much for your willingness to participate! We understand that under the current circumstances, your time is extremely limited. That makes us all the more grateful for your enthusiasm and commitment.

Best wishes,

The Q-BEx Team

## **Supplement 6**

**Paper title:** How to Quantify Bilingual Experience? Findings from a Delphi Consensus Survey

**Content:** A list of themes and their definitions used to code the data in the round 1 thematic analysis



## Round 1 Thematic Analysis

### Themes:

**Alternatives** – any other solution or proposal of how to document a construct. For example, if we suggest documenting past exposure/use year-by-year, and the panellists disagree with this and suggest measuring it based on certain occasions in child's life that led to changes in language exposure/use (e.g. when the child moved to a different country, when they started school, etc.). Note: this theme should be differentiated from 'in relation to a specific variable' (see below) or 'change in the number of categories' (see below).

**Atypical populations** – comments in relation to the tool/construct and any participants with non-neurotypical development or the diagnosis. For instance, when we suggest that we need a common tool to measure bilingual experience, and the panellists agree and comment that such a tool would help differentiate between children with typical and atypical development. In a sense, such comments could be covered with the theme 'flexibility' (meaning that the tool should be flexible enough to be used both with typical and atypical populations), but it might be better to flag anything to do with atypical development with a separate theme.

**Change in the number of categories** – this theme is somewhat more narrow than the theme 'alternatives'. When a statement suggests documenting a specific construct through a set of categories (or one category), this theme would apply in the following cases: (a) suggestions of additional categories; (b) suggestions of removing some of the listed categories; (c) suggestions of regrouping or merging categories. For example, if we suggest measuring language use with (1) parents, (2) siblings, and (3) friends, but the panellists suggest one of the following: adding more interlocutors, excluding friends, or merging parents and siblings under the category 'family'. Even though this is an alternative solution to our proposal, we wouldn't use the theme 'alternatives' but rather 'change in the number of categories' to be more specific.

**Clarity** – anything relating to the panellists or future participants or tool users not understanding the statement/tool because of its wording, ambiguity, vagueness. For instance, when we say that the questionnaire would be applicable to children who speak more than two languages, and the panellists suggest rewording this into 'two or more than two languages'. Also this theme can be used when the panellists find the statement confusing for any reason or when a specific concept is vaguely defined or it needs to be defined (for instance, when we say that the frequency of use should be documented separately for each type of interlocutor and the panellists wonder about the meaning of the phrase 'type of interlocutor'). This theme might be the most important to consider, as we want to make sure that we don't miss out the statements that were not clear in round 1.

**Consistency** – any comments in relation to documenting specific constructs in a consistent way or comments that mention that one same construct is (not) referred to in the same way across the disciplines, labs, etc. For instance, when we suggest measuring language exposure in school in a specific way, and the panellists agree, but they comment that this should be done in a comparable way to documenting language exposure in the home. Or

when the panellists comment that what some researchers consider language quality is called something else by a different research group.

**Ethics** – anything to do with ethical issues, informed consent, etc. For instance, when suggesting that we should measure language history, and the panellists comment that this might require informed consent and that the procedure might vary from country to country.

**Flexibility** – comments in relation to the optionality of specific parts or the ability to adapt to specific context-based needs, participant-based needs, or researcher/practitioner-based needs. For instance, when we suggest documenting language exposure year-by-year, and the panellists agree but comment that we do not require the same level of detail for every year. Or for instance, when the panellists agree that we need a common set of measures to document bilingual experience, but that this should be adaptable to various environments or groups. Note: this should be differentiated from the theme 'atypical populations'.

**Follow-up** – suggestions about future work on the questionnaire or in the field. For instance, an in-depth case study might help us determine if we should document a specific construct in the way that we propose.

**In relation to a specific variable** – when the proposed variable/construct should be documented/considered in relation to another variable or is related to another variable. For instance, when we suggest documenting exposure/use year-by-year, and the panellists agree but they add that this should be done only in younger children, but not in older kids, or only in children from a specific SES background. In a sense, this is not a fully different proposal, so we do not label this comment with the theme 'alternatives'. Rather, the panellists suggest agreement with documenting a specific variable in a year-by-year manner, but in relation to age or SES. Therefore, the theme 'in relation to a specific variable' is more adequate.

**Misconceptions** – anything to do with the misunderstanding or wrong assumptions about the concepts, contexts, measures, etc. For instance, when we suggest documenting overheard speech by parents and the panellists suggest that this information should be taken with a pinch of salt. Or when we suggest documenting language use during holidays and the panellists comment that traveling to a home country does not necessarily imply using the home language.

**Misuse** – anything to do with the misuse of the tool. This would refer to comments such as 'we risk using measures for bilinguals inappropriately with multilinguals in the same way that sometimes we use measures intended for monolinguals with bilingual children'. Note: this theme could overlap to an extent with 'misconceptions' or 'prescriptivism'.

**Open science** – anything to do with shareability and making the tool available between sectors.

**Opt-out** – expressions of the inability of panellists to comment due to the lack of knowledge/expertise, or because they can't make up their mind, but they don't provide a

reason why, which would be covered by any other theme. For instance, when we propose documenting digital language exposure/use, and the panellists comment 'not so sure about this'. Note: this could slightly overlap with the theme 'requirement in practice' (see below). However, the theme 'opt-out' is preferred when the panellist expresses uncertainty or lacks knowledge to make a judgement.

**Overlap** – any indications that certain proposed measures/constructs overlap in what they measure/estimate. For instance, this would refer to the comments pointing out that if we document language use at home, and language use at school, and then also digital language use, there might be some overlap as the digital language use probably happens in the home and/or at school. In that case, we would be estimating the same thing twice.

**Practicality** – anything related to the feasibility of doing something. For instance, when the panellists agree that we need to document overheard speech or that we need to document language mixing, but they express concerns that this might not be possible to do (in an objective way).

**Prescriptivism** – anything to do with this project imposing rules across the fields. For instance, even if we agree that there should be a common tool used across sectors, this should not prevent other tools from being proposed or created.

**Requirement in practice** – necessary or not necessary within or across the sectors. For instance, this theme would be adequate when the panellists agree that a specific construct needs (or doesn't need) to be documented, or that the comparability of data is required. Note: even though the word 'practice' is used in the name of the theme, this applies to all stakeholder groups, not just to practitioners.

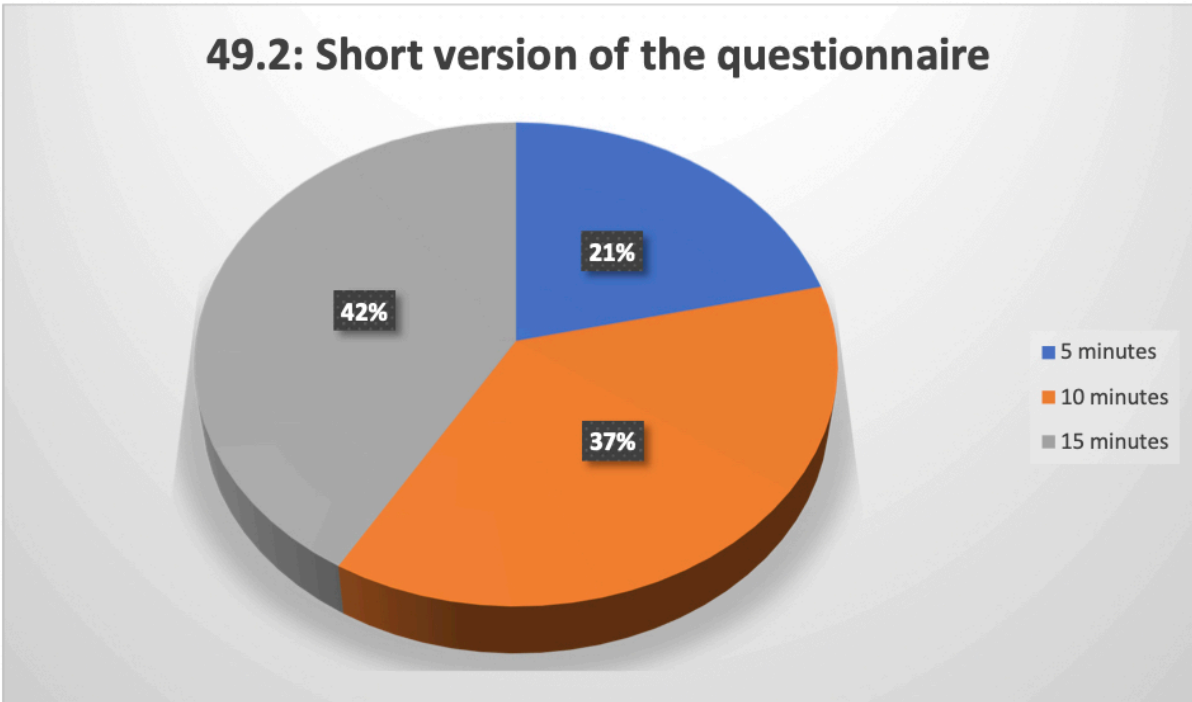
**Stating related facts** – comments stating facts, beliefs or common features of one's practice or of the field. For instance, 'I have a lot of bilinguals on my caseload' or 'bilingualism is a dynamic process'. These comments do not seem to have consequences on our aims; we just used this thematic label to mark them somehow.

**Other** - anything else not covered by the above themes.

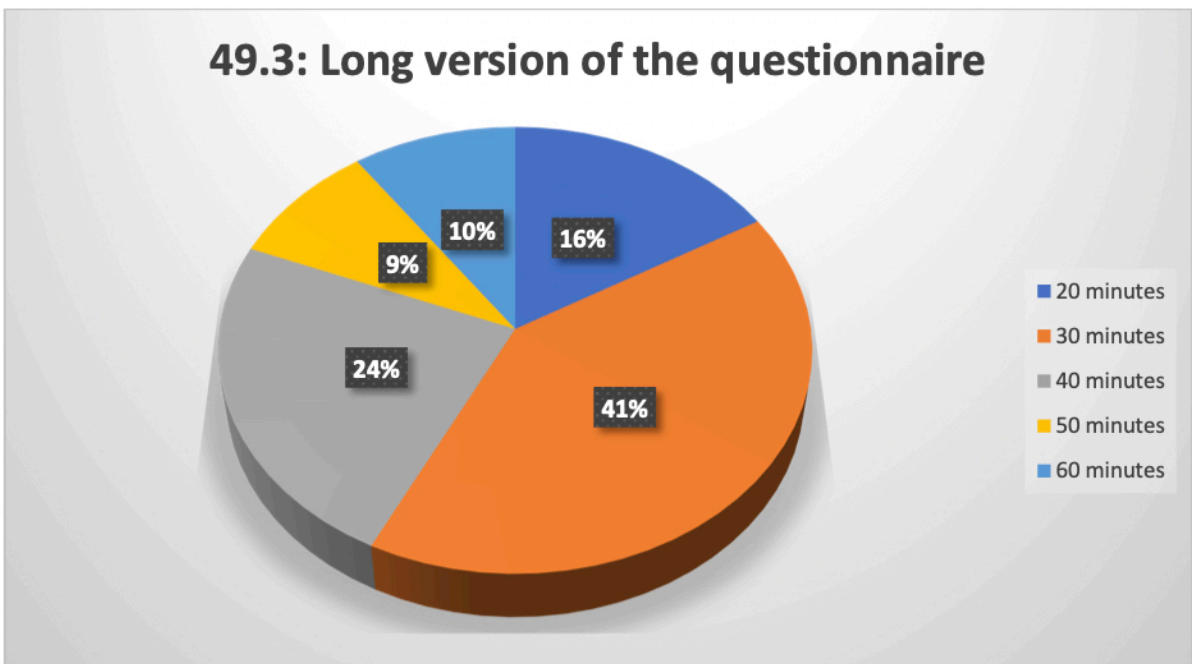
## **Supplement 7**

**Paper title:** How to Quantify Bilingual Experience? Findings from a Delphi Consensus Survey

**Content:** Distributions of responses on the two questions inquiring about the preferred length of the short and the long versions of the questionnaire (statements 49.2 and 49.3 respectively).



**Fig. S1** Distribution of responses on statement 49.2 inquiring about the length of the short version of the questionnaire

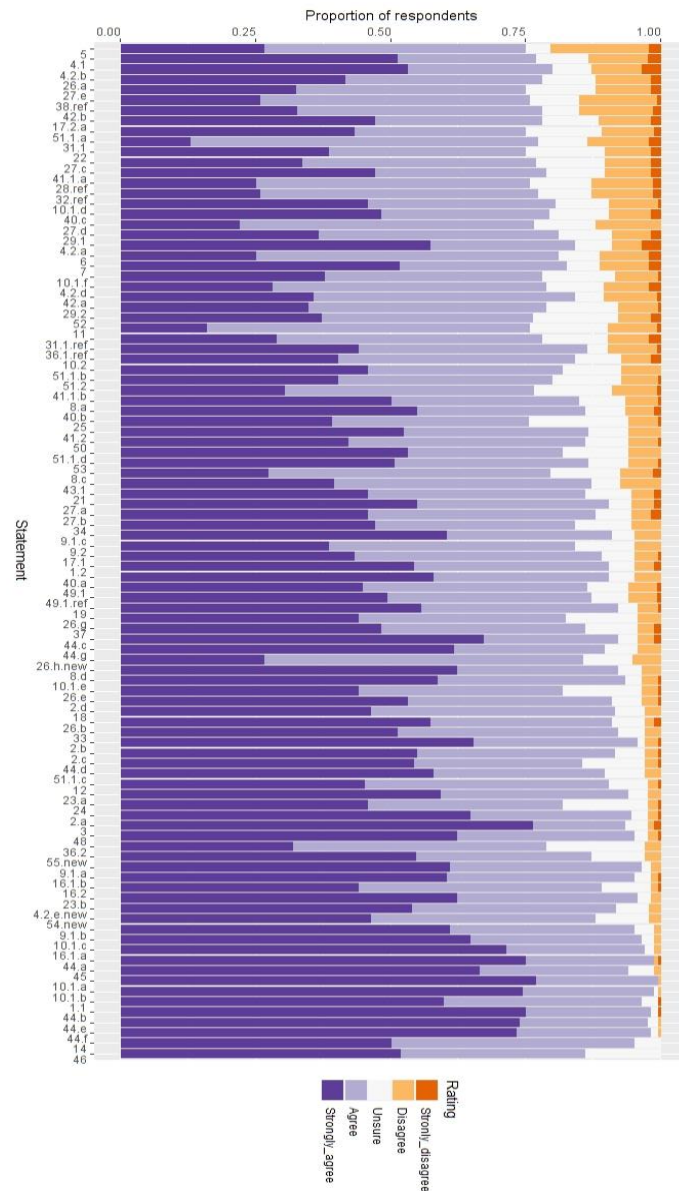


**Fig. S2** Distribution of responses on statement 49.3 inquiring about the length of the long version of the questionnaire

## **Supplement 8**

**Paper title:** How to Quantify Bilingual Experience? Findings from a Delphi Consensus Survey

**Content:** Distribution of ratings among consensus-reaching statements, ordered by level of disagreement.



**Fig. S3** Distribution of ratings among consensus-reaching statements, ordered by level of disagreement.