# *Q methodology in applied linguistics: a systematic research synthesis*

System 120 (2024) 103194

Contents lists available at ScienceDirect

System

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

# Q methodology in applied linguistics: A systematic research synthesis

Nicola Morea [a,*], Hessameddin Ghanbar [b]

[a] Institute of Education, University of Reading, United Kingdom
[b] Islamic Azad University, Fereshtegaan International branch, Iran

ARTICLE INFO

*Keywords:*
Q methodology
Q-sort
Mixed methods
Systematic review
Research synthesis
Applied linguistics
Language education

ABSTRACT

Q methodology (Q) is a mixed-methods research methodology used to systematically explore people's subjectivity. Despite an increasing number of theoretical and empirical studies in applied linguistics (AL) adopting this methodology, we found a lack of systematic research syntheses examining how Q has so far served the field, especially considering the complexity of Q in relation to instrument creation, data collection and data analysis. To address this gap, we conducted a database search to map the uses of Q in AL research, leading to the identification of 55 empirical studies. An analysis of the contextual, methodological and data-analytical characteristics of the selected studies showed that the Q-sort method is being increasingly used in AL as a tool to encourage participant reflexivity, particularly in the areas of teacher and learner cognition and emotions, and language-specific and multilingual motivation. However, the review also revealed current gaps in the applications of Q, and most notably a frequent lack of quality-assurance practices during the development of the Q-set and omissions of important data-analytical information in published research, reducing the transparency and replicability of findings generated from Q studies. We conclude with recommendations for future research using Q in AL.

## 1. Research background

### 1.1. Introduction

Q methodology (Q) is a research methodology designed to systematically examine people's perspectives on complex and subjective matters. Whilst Q originated in psychology, it has been employed in other research fields, such as health sciences, political science and education. Q is also beginning to emerge in applied linguistics (AL), as evidenced by several conceptual papers introducing the methodology to the field (e.g., Irie et al., 2018; Li, 2022; Thumvichit, 2022a) as well as a growing number of empirical studies that have adopted Q to explore people's viewpoints and experiences around language-related phenomena. The increasing popularity of this methodology is also evidenced by the emergence of research syntheses examining the applications of Q in various fields, such as education (Lundberg, de Leeuw, & Aliani, 2020), healthcare (Churruca et al., 2021) and nursing education (Hensel, Toronto, Lawless, & Burgess, 2022), or across research fields (Dieteren, Patty, Reckers-Droog, & van Exel, 2023; Morea, 2022). To date, however, there does not seem to be any systematic research synthesis of the use of Q methodology in AL. The growing interest in Q from applied linguists, coupled with the complexity of this methodology, indicates that the time is ripe for such an exploration. Accordingly, a

* Corresponding author. Institute of Education, University of Reading, London Road Campus, Reading, RG1 5EX, United Kingdom.
*E-mail address:* n.morea@reading.ac.uk (N. Morea).

https://doi.org/10.1016/j.system.2023.103194
Received 4 September 2023; Received in revised form 1 December 2023; Accepted 1 December 2023

systematic review of the literature was conducted, with the aim of mapping the applications of Q within AL research and reviewing the contextual, methodological and data-analytical characteristics of Q studies in the field. Since Q represents a relatively new methodology in AL, we first introduce and describe its most salient features before discussing the approach used for selecting and reviewing the literature.

### 1.2. Principles of Q methodology

Q was invented by physicist and psychologist William Stephenson in the 1930s. Stephenson's aim was to provide an alternative to the dominant approach of measuring psychological traits and studying individual differences through scales and tests administered to large samples from a target population (Stephenson, 1953). Instead, Stephenson proposed a revised application of factor analysis which could be applied to a small number of people (or even single cases) to bring the subjectivity of the individual to the fore. This approach, referred to as the inverted factor technique or by-person factor analysis (Watts & Stenner, 2012), allows the researcher to group a relatively small number of participants (instead of a sequence of variables, as in traditional factor analysis) based on how similarly they scored a large number of items. The items scored by participants in a Q study take the form of a *Q-sort*. A Q-sort is the result of a sorting activity, in which participants rank-order a set of items (commonly written statements) in a way that reflects their opinions and/or experiences on a topic of inquiry.

In Q, a person's Q-sort is considered a manifestation of their subjectivity. When multiple Q-sorts are collected from a sample of participants, these can be analysed using by-person factor analysis to detect common configurations of viewpoints (i.e., factors) within a sample. Then, participants' configurations of viewpoints can be described and contrasted by qualitatively analysing each factor. This combined use of quantitative and qualitative analyses into a single analytical procedure is what makes Q an "inherently mixed methods approach" (Lundberg et al., 2020, p. 1; Ramlo, 2016).

The Q procedure summarised above can be divided into four key stages. In the next sections, each stage is explained more in detail, as to provide a reader not familiar with Q with an understanding of its key steps.

#### 1.2.1. The design stage: creating the Q-set

The creation of the Q-set, namely the set of items that participants are asked to rank-order into a Q-sort, follows a procedure specific to Q. Just like research participants represent a sample from a target population, the Q-set items represent a sample of statements from a *concourse* of reference. Brown (1993) defines the concourse as "the flow of communicability surrounding any topic" (p. 94). Items from a concourse can be accessed by interviewing people who are familiar with the topic of inquiry to explore their perspectives, or indirectly by reviewing various sources, from academic literature to newspapers and magazines (Brown, 1993). After obtaining a large number of items from the concourse, the researcher reduces them into a manageable set. This can be done by removing repetitive items and by ensuring that the resulting Q-set is both balanced and diverse in terms of the opinions expressed (Watts & Stenner, 2012).

#### 1.2.2. The data-collection stage: Q-sorts and post-sorting interviews

Once the Q-set has been designed and refined, research participants are recruited. In Q, the participant-sampling process is purposive, and the researcher should recruit participants who are sufficiently familiar with the topic of the study and thus able to sort the Q-set meaningfully (Watts & Stenner, 2012). Although it is common for Q studies to involve a group of participants, single-case Q studies, involving the creation of several Q-sorts from the same participant either over time or with different conditions of instruction, also exist.

The data-collection stage primarily involves the generation of Q-sorts. Each participant is asked to rank the Q-set items onto a predefined grid, (typically shaped as a quasi-normal distribution, see Fig. 1) according to a condition of instruction (e.g., degree of agreement). As a result of this forced distribution, there is a fixed number of items that participants can place in each column of the

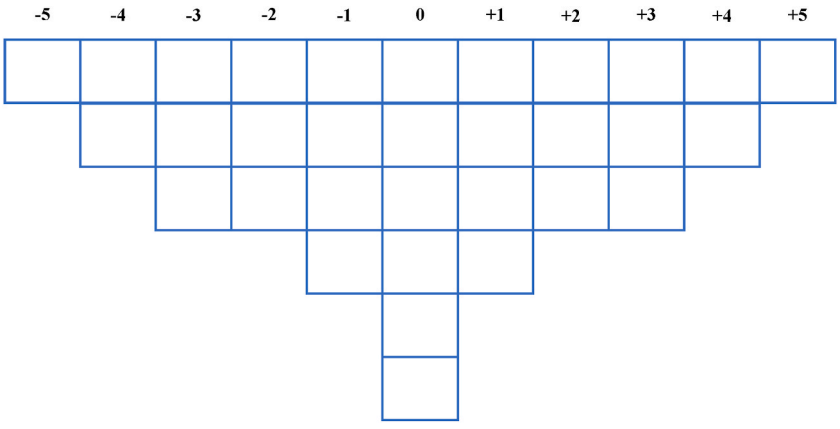


**Fig. 1.** A Q-Sort grid containing 32 Q-items.

grid, with fewer items that can be placed at the extremes of the grid and more items towards the middle. This encourages participants to consider the items in relation to each other rather than in isolation, and to prioritise some items over others. Once all the items have been placed on the grid, a participant's Q-sort is complete. The researcher then interviews participants to gather insights into the reasons behind participants' sorting decisions.

Whilst the Q data-collection stage has been traditionally conducted in person, with items being printed and cut into a set of cards, online Q studies are on the rise, involving the use of specialised websites and pieces of software to generate online Q-sorts (see Alanazi, Wharrad, Moffatt, Taylor, & Ladan, 2021; Meehan, Ginart, & Ormerod, 2022, for more information on how to design and conduct an online Q study).

#### *1.2.3. The data-analysis stage: Q-sort correlation and Q factor analysis*

The analysis of Q-sorts is conducted on specific programmes for Q factor analysis, such as PQ-Method (Schmolck, 2014). The software first correlates each participant's Q-sort with the Q-sorts of all other participants, producing a correlation matrix indicating the degree of similarity between pairs of Q-sorts. As in traditional factor analysis, the successive steps involve a series of decisions regarding the methods and criteria to be used for factor extraction, retention and rotation. After specifying a factor-extraction method (e.g., the centroid method), the software extracts a number of factors and provides additional information (e.g., eigenvalue, explained variance, scree plot). The researcher selects a factor solution (i.e., which and how many factors to retain) based on specific criteria (e.g., minimum factor eigenvalue, overall explained variance, examination of the scree plot) and rotates the selected factors. Although Q methodologists have traditionally favoured judgmental or theoretical rotation (whereby the researcher manually rotates the factors until a satisfactory factor solution is achieved), the use of statistical rotations like varimax seems equally widespread (see Akhtar--Danesh, 2017, for an extended discussion on Q factor extraction and rotation techniques).

Once the retained factors have been rotated and a final solution obtained, the software produces a table of factor loadings, namely correlation coefficients indicating how close each participant's Q-sort is to each retained factor. At this stage, the researcher *flags* (i.e., selects) the Q-sorts that will be used for calculating factor scores. Only Q-sorts whose factor loading on a factor is statistically significant (at the $p < .05$ or .01 level) are typically selected. The software then uses the factor scores obtained to generate a factor-specific Q-sort called *factor array*. The factor array appears like any other Q-sort, but the disposition of the items on the grid is based on each item's factor score (i.e., a weighted average of the value assigned to the item across the Q-sorts significantly loading on a given factor). The factor array can thus be considered an archetypical, factor-defining Q-sort.

#### *1.2.4. The data-interpretation stage: understanding the retained factors*

The final step requires the researcher to examine the factor arrays to interpret each factor, establishing what makes each factor unique and different from the others. Whilst the previous stages of data analysis are based on quantitative procedures, the interpretation of factors is a purely qualitative endeavour. The researcher examines the configuration of items in the factor array both within and between factors. This is a holistic process by which the researcher should consider the whole configuration of items (rather than only focusing on the items appearing at the extremes of the factor-array grids), together with other qualitative data generated from participant interviews.

### *1.3. Potentials and limitations of Q*

Various advantages and limitations have been attributed to Q. More generally, Morea (2022) stresses that Q favours a holistic investigation of people's subjectivity. This applies both to the sorting activity, since items are considered in combination rather than in isolation (as opposed to Likert items), and to the data interpretation process, given that the researcher examines and describes each factor in its entirety rather than focusing on individual items. Through its cognitively demanding data-collection process, which requires an active engagement from participants both when comparing and ranking the Q-items and when explaining the sorting decisions, Q has the potential to provide a more in-depth investigation of individual's subjectivity than traditional methods like questionnaires. When considering the potential of Q for AL research specifically, both Irie et al.(2018) and Thumvichit (2022a) argue that Q can be applied to a variety of areas in AL and second language (L2) research, such as in investigations into language users' emotions, beliefs, identity, agency, motivation and values in relation to language-related phenomena. Additionally, Irie et al.(2018) and Lundberg et al. (2020) stress that a quality of Q is its participatory approach, due to participants' "active engagement in the research task" (Irie et al., 2018, p. 28), with Lundberg et al. (2020) arguing that Q has the potential to bring the voices of the "otherwise marginalized" to the fore (Lundberg et al., 2020, p. 12, but also Brown, 2006). The potential of Q for participatory research is highlighted by the common practice of conducting post-sorting interviews, where participants are encouraged to reflect on the sorting process and explain their decisions, to the extent that the Q-sort activity could also be used as an educational activity to promote autonomy and self-awareness (Irie et al., 2018; Pemberton & Cooker, 2012).

There are also limitations that should be equally considered. Firstly, as highlighted in Lundberg et al. (2020), the creation of a Q-set is a time-consuming and iterative process, especially if the Q-items are obtained after interviewing members of the target population. Furthermore, the lack of shared and structured practices for drawing statements from the concourse and refining them into a Q-set may pose the risk of researcher bias, especially if the process of Q-set development does not involve any form of quality assessment. Finally, an important aspect to highlight regards the generalisability of the results of a Q study. Whilst it has been argued that Q does not allow the researcher to make generalisations from the research findings due to the small number of participants and its purposive sampling approach (Li, 2022), this point requires a clarification. Q is certainly not concerned with making inferences on neither the frequency of the viewpoints revealed in the target population (e.g., the factor explaining most of the sample variance may not be the most frequent

viewpoint held within the target population), nor with testing whether individual traits (e.g., age, socio-economic status, gender) may explain participants' association with a specific factor. However, as argued by Stephenson (1953), the researcher can be confident that the factors emerging from the sample of participants represent shared configurations of viewpoints that also exist in the population of reference.

### *1.4. Methodological syntheses in AL: making a case for synthesising Q research in the field*

Synthesising research practices has gained momentum in AL and L2 research recently. These research syntheses have encompassed a wide variety of areas, ranging from scale quality in different domains such as anxiety, motivation, and willingness to communicate (Sudina, 2021, 2023), task-based language learning and teaching (Bryfonski & McKay, 2019; Plonsky & Kim, 2016), assessment tasks like grammaticality judgment tests (Plonsky, Marsden, Crowther, Gass, & Spinner, 2020) and elicited imitation tasks (Kostromitina & Plonsky, 2022; Yan, Maeda, Lv, & Ginther, 2016) to proficiency assessment techniques in L2 studies (Park, Solon, Dehghan-Chaleshtori, & Ghanbar, 2022; Tremblay, 2011) and L2 writing (Zhang, Gibbons, & Li, 2021; Zhang & Plonsky, 2020). In relation to research syntheses more relevant to this work, we have been witnessing a burgeoning number of methodological syntheses in AL generally, and in L2 research specifically, focusing on exploratory factor analysis (Plonsky & Gonulal, 2015), multiple regression (Plonsky & Ghanbar, 2018), structural equation modelling (Ghanbar & Rezvani, 2023; In'nami & Koizumi, 2011), Rasch measurement (Aryadoust, Ng, & Sayama, 2021), cluster analysis (Crowther, Kim, Lee, Lim, & Loewen, 2021) and meta-analyses in second language research (Vuogan & Li, 2023). Nonetheless, relating to Q, our searches revealed a dearth of methodological syntheses, although we have seen some non-systematic reviews of the literature likes notes or commentaries. For example, Thumvichit (2022a) and Li (2022) introduced Q to L2 researchers by providing a historical background and expounding upon its related theoretical concepts and issues as well as showcasing a study to exemplify how Q might be used in L2 research. In a similar vein, and approximately nine years beforehand, Irie et al.(2018) also introduced Q to the field by using a case exemplar. Ultimately, although these studies were about Q and showcased its use, thus providing valuable methodological information, they did not involve any systematic and formalised data-collection and data-analysis stages, and thus cannot be considered research syntheses of the uses of Q methodology in AL research (see Chong & Plonsky, 2023, for a discussion of the requirements of systematic reviews). It should also be mentioned that notwithstanding the fact that Morea (2022) and Lundberg et al. (2020) recently conducted two research syntheses about Q, the former focused on longitudinal studies (observational and experimental) across research fields and the latter reviewed compulsory-education research published between 2010 and 2019.

Considering the gap in our knowledge around the extent to which Q has been used in AL, we consider timely and necessary to conduct a methodological synthesis of Q studies and evaluate to what extent the recommendations provided in the aforementioned papers align with current research in the field. Importantly, the methodological and data-analytical stages of Q involve a considerable number of decisions (e.g., decisions around Q-set creation, factor selection, retention and rotation methods), inevitably resulting in variability in research practices. A systematic review can reveal the most common research practices currently used in the field and lead to informed recommendations for rigorous future applications of this methodology.

### *1.5. Study aims*

Following Chong and Plonsky's (2023) classification of research syntheses in AL, this research can be considered a methodological synthesis on the uses of Q in AL research. Rather than focusing on the findings generated from research using Q, this review is aimed at mapping the applications of Q in AL, together with examining the methodological characteristics of the reviewed studies to provide a reference for applied linguists interested in Q. As a result, this review was guided by the following overarching research questions (RQ).

**RQ1.** To what extent has Q methodology been used so far to explore language-related issues and phenomena?

**RQ2.** What are the Q-specific characteristics of the selected studies?

## 2. Method

### *2.1. Study retrieval and identification*

Guided by other systematic reviews in the field (e.g., Park et al., 2022; Riazi, Shi, & Haggerty, 2018) and by Plonsky's (2013, 2014) criteria for outlining the domain of a study (i.e., substantive, locational, temporal), we first selected the inclusion criteria of this research synthesis (the substantive dimension). The following criteria were used to identify studies to be included in the review:

- The study must be empirical;
- The study must investigate a language-related phenomenon;
- The study must employ Q methodology and use its key methodological features, namely Q-sorting and Q factor analysis.

The first selection criterion implied the exclusion of conceptual papers introducing Q methodology without involving any empirical investigation. Regarding the second inclusion criterion, we decided to maintain a broad approach in delimiting the field of AL. Our key concern was that the studies included in the review were exploring a language-related issue or phenomenon as at least one of their

primary research aims. Therefore, studies in which language represented a related but secondary objective (e.g., studies on intercultural competence) were excluded. The third inclusion criterion aligns with the approach used by Lundberg et al. (2020) in their systematic review of Q research in the field of compulsory education. This criterion allowed us to exclude studies that, whilst adopting the Q-sort method, did not align with the fundamental procedural principles of Q methodology. Finally, only certain document types were considered for inclusion, namely journal articles, theses and dissertations, books, and book chapters.

The study identification process is displayed in Fig. 2 via a PRISMA flowchart (Page et al., 2021). Regarding the locational dimension, the team of two researchers consulted the following databases at the end of July 2022: ERIC, LLBA (Linguistics and Language Behavior Abstracts), PsycINFO, DOAJ (Directory of Open Access Journals), ProQuest Dissertation and Theses, ScienceDirect Journals, JSTOR Arts and Sciences, SAGE Journals Premier, Wiley Online Library, Web of Science (ISI/Thomson), and Scopus (Elsevier). The search terms used were "Q method*" OR "Q-method*". The use of the wildcard character allowed us to search for variations (e.g., "Q methodology" and "Q methodological", with and without a hyphen). When a database was not specific to language-related research, relevant domains were selected. For example, when searching the Scopus database, the following fields were specified: "Social Sciences", "Psychology", "Arts & Humanities", "Neurosciences" and "Multidisciplinary". Finally, regarding the temporal dimension of this study, it should be mentioned that since one of the aims of this review is to map the applications of Q in the field over time, no timeframe was specified.

With regard to study identification, we first screened the titles, abstracts and key words of the records obtained from each database search. After removal of duplicate records, 82 potentially eligible studies were identified. The 82 studies were divided between the two researchers and their full text examined against the inclusion criteria. Studies for which the inclusion decision was not straightforward were examined by both researchers, until a common decision was reached. Three additional records were added via snowballing, as they were known by the authors but did not appear in the database search. In line with the other studies, these were all published no later than July 2022. The number of records that were examined in their entirety was thus 85. Of these, 30 records were removed for the following reasons: the study did not report new empirical findings (n = 10); the full text of the study could not be obtained (n = 8); the study did not use Q or conducted Q factor analysis (n = 4); the full text was in a language not known to the researchers (n = 3); the study did not investigate a language-related phenomenon (n = 3); and the study was not in one of the formats specified for inclusion (n = 2, both meeting papers). The final number of records included in this review was thus 55. The full reference of these studies can be found as supplementary material.

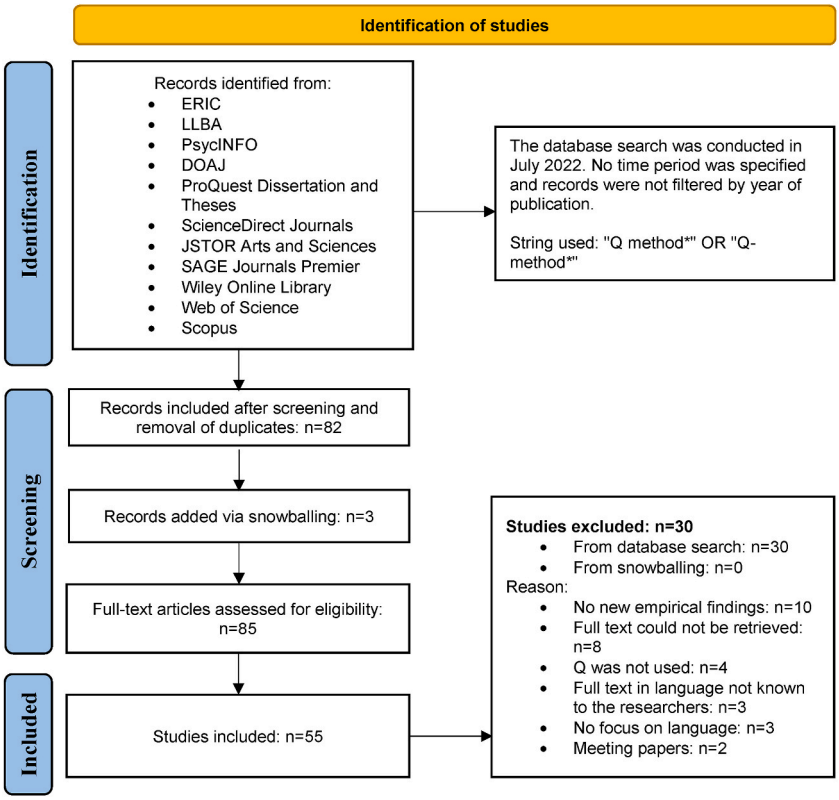


**Fig. 2.** Study-identification process.

### 2.2. Coding

The coding scheme used was informed by the systematic reviews of Q research by Lundberg et al. (2020) and Morea (2022), as well as by the conceptual papers on Q in AL by Irie et al.(2018), Li (2022) and Thumvichit (2022a). The coding scheme was made of two sections. The first section consisted of variables related to the context and topic of the studies, and namely: 1. publication type (e.g., journal article, book chapter, dissertation), 2. year of publication, 3. study domain, 4. study focus, and 5. research design (i.e., cross-sectional, longitudinal, single-case study). The categories used to group the selected studies based on their domains were created after a qualitative interpretation of the context and research objectives of each selected study. More precisely, we used bottom-up thematic analysis (Braun & Clarke, 2006) in order to find major and recurrent themes in annotations we gathered relating to the domain of studies. We also used a peer-debriefing process (Creswell, 2003) to check the trustworthiness of data analysis in this part. Regarding the foci of the selected studies, we followed Lundberg et al.'s (2020) categorisation. After inspecting the research questions and aims, as well as the condition of instructions of the sorting activity, we assigned each study to one of the following categories (from Lundberg et al., 2020):

- Representation (Q was used to explore "participants' representation of a subject matter" and to understand "how the issue under scrutiny is typically understood" by the research participants, p. 6).
- Attitudes and values (focusing on participants' preference in relation to the issue being investigated).
- Critical reflection (participants were asked to "critically reflect about their situation or characteristics", p. 6).
- Evaluation (participants were asked to evaluate a specific issue).
- Response (focusing on "participants' preferences in terms of how to respond to a certain subject matter", p. 7).
- Decision making (focusing on participants' decision-making process).
- Multiple (two or more of the above categories could be assigned to the study).

The second section of the coding scheme focused on the methodological characteristics of the studies, and it was subdivided into the following four areas: 1. Q-sample characteristics, 2. P-sample characteristics, 3. Q-sorting characteristics, and 4. data-analysis characteristics. Q-sample characteristics refers to the Q-set design process and the form of the Q-items used. This area was comprised of the following variables: 1. sources from which the Q-items were drawn (e.g., academic literature, policy documents, newspapers), 2. quality procedures used when designing the Q-set (e.g., piloting, expert evaluation), 3. Q-set size, 4. form of the Q-set (e.g., written, pictorial), and 5. type of Q-items. Regarding the latter variable, Thumvichit (2022a) explains that, based on their form and purpose, Q-items can "take the form of questions, opinions, actions, experiences, plans, strategies or situations" (p. 6), and our coding reflected this classification. Next, P-sample characteristics pertains to the participant characteristics, and encompassed the following variables: 1. participant sample size, 2. national context, 3. setting (e.g., schools, universities), and 4. participant role (e.g., teacher, learner). The area named Q-sort characteristics refers to how the sorting activity was conducted. The following variables were considered in this part of the coding scheme: 1. scale range of the Q-sort grid, 2. use of forced distribution, 3. modality (e.g., in-person, online), 4. software used for online Q-sorting, and 5. whether post-sorting interviews were conducted. Finally, data-analysis characteristics comprised of the following variables: 1. software used to conduct by-person factor analysis; approach used for 2. factor extraction, 3. factor retention and 4. factor rotation; 5. number of retained factors; 6. Q-sort flagging criteria; 7. treatment of confounded Q-sorts; 8. total variance explained by the factor solution; and 9. factor interpretation (i.e., whether each retained factor was interpreted and described).

To pilot our coding approach and calculate intercoder agreement, we selected ten studies (18% of the total) that both researchers coded. We estimated the percentage of agreement for each variable of the coding scheme and the initial mean of agreement percent was 96%. When differences in coding were found, these were discussed until an agreement was reached. Specifically, disagreement was resolved by adding a "multiple" category to the "study focus" variable. Similarly, regarding the second section of the coding scheme, a "multiple" category was added to the "type of Q-items" variable. No disagreement was found in relation to other variables. The coding scheme can be found in the IRIS database (Marsden, Mackey, & Plonsky, 2016), and it can be accessed through the following link: https://www.iris-database.org/details/zllKb-m4qCM.

### 2.3. Data analysis

To address the first RQ, a scholarly work was deemed as the unit of analysis. Following this, we use frequencies and percentages to map the frequency and type of Q studies in the field over time, together with the domains and foci of studies. We also utilise descriptive statistics like frequencies and percentages to report the results of the second RQ, and for continuously measured variables (e.g., Q-set size, variance explained and number of retained factors), we report mean and standard deviations. Percentages are reported as whole numbers after rounding decimals (using .5 as the minimum threshold for rounding up); hence, some totals may be lower than or exceed 100%. Importantly, for the variables related to the second RQ, *study components* were used as units of analysis. In line with Lundberg et al. (2020), we considered a study to have multiple components when it involved the generation of more than one set of Q-sorts, either by administering a different Q-set or by administering the same Q-set at successive time points (for example after an intervention). Whilst most of the studies only featured a single component, nine of the 55 reviewed studies involved the generation of multiple Q-sorts from the same participants, either at the same time point (but with a different Q-set and/or condition of instruction, n = 5) or at multiple time points (using the same Q-set and condition of instruction, n = 4). Since these studies involved multiple data collections and analyses, Q-specific characteristics like Q and P-sample size, participant identity and analytical procedures could differ within the same study. Therefore, a new entry (i.e., an additional row in the codebook) was created for each component within those

studies, resulting in a sample size of n = 65 (n = 8 studies consisted of two components and n = 1 consisted of three).

## 3. Results

### 3.1. RQ1: mapping the uses of Q in AL research

#### 3.1.1. Publication type and spread over time

A breakdown of the 55 selected studies by publication type showed that most studies were journal articles (n = 43, 78%), followed by theses and dissertations (n = 9, 16%) and books and book chapters (n = 3, 5%). In terms of publication year, the first Q study in the field that we identified was published in 1999 (Boscolo & Cisotto, 1999). From that time, the number of Q studies in the field has been soaring, particularly since 2019 (Fig. 3). For example, only in the first half of 2022 (the review was conducted in July 2022), we could find 11 pieces of published research.

#### 3.1.2. Domains and foci

After a thematic analysis of the study topics and aims, we could group the 55 selected studies into nine themes reflecting different research domains within AL research (Fig. 4):

1. **Teacher cognition and emotions** (n = 22, 40%): these studies either focused on pre-service and in-service teachers' beliefs (and in one case, parental beliefs) about a variety of issues (e.g., literacy instruction, multilingualism) or on teachers' emotions (e.g., anxiety). Whilst being the most common study domain, these studies are also relatively recent, as evidenced by their publication year (2011–2022), indicating the noticeable research attention around this area from applied linguists and Q methodologists.
2. **Learner cognition and emotions** (n = 10, 18%): these studies explored either learners' viewpoints about specific aspects of language learning (e.g., the role of grammar in English for Academic Purposes, specific learning activities during TESOL instruction) or their emotions in relation to language learning (e.g., boredom). These studies were published between 2004 and 2022, suggesting that Q research in AL on language learners has preceded the currently more widespread focus on teachers.
3. **Language-specific and multilingual motivation** (n = 9, 16%): these studies tended to investigate language learners' (and in two cases language teachers') motivation, framed as either language-specific (i.e., motivation to acquire a specific additional language) or multilingual. This also represents a relatively new research domain for studies using Q, with publication year spanning from 2014 to 2022.
4. **Literacy education** (n = 3, 5%, publication year: 1999, 2013, 2014) (e.g., strategies for teaching to write in primary/elementary schools).
5. **Educational language policy and planning** (n = 2, 4%, publication year: 2020, 2021) (e.g., stakeholders' perspectives on a university's educational language policy).
6. **Additional language pedagogy** (n = 2, 4%, publication year: 2012, 2013) (e.g., modes of language learners' autonomy).
7. **Language disorders and deaf studies** (n = 4, 7%, publication year: 2001, 2007, 2010, 2020): (e.g., perspectives on the therapeutic alliance for people with aphasia, perceptual characteristics of voice-hallucination in deaf people).
8. **Sociolinguistics issues** (n = 2, 4%, publication year: 2020, 2021) (e.g., Korean citizens' perspectives on the meaning of a Korean word).
9. **Translation studies** (n = 1, 2%, publication year: 2019) (i.e., views on enhancing legal translation competence).

It should be highlighted that studies in categories one to six share a strong educational element, indicating the prevalence of studies at the intersection between AL and education in our sample (n = 48, 87% when combining categories 1–6).

Regarding the foci of our selected Q studies (see explanation of the categories used in Section 2.2), *critical reflection* was the most dominant focus (n = 16, 29%) followed by *representation* (n = 15, 27%), and *response* (n = 8, 15%) (Fig. 5). The category *multiple* (n = 7, 13%) was comprised of: (a) representation and response (n = 5), (b) critical reflection and response (n = 1) and (c) representation and

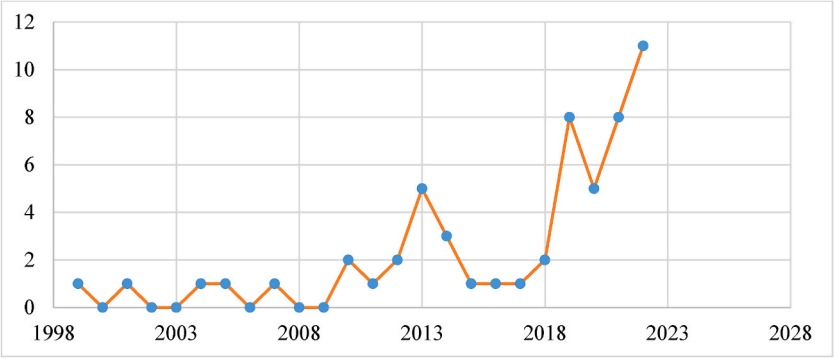


**Fig. 3.** Publication year of selected studies.

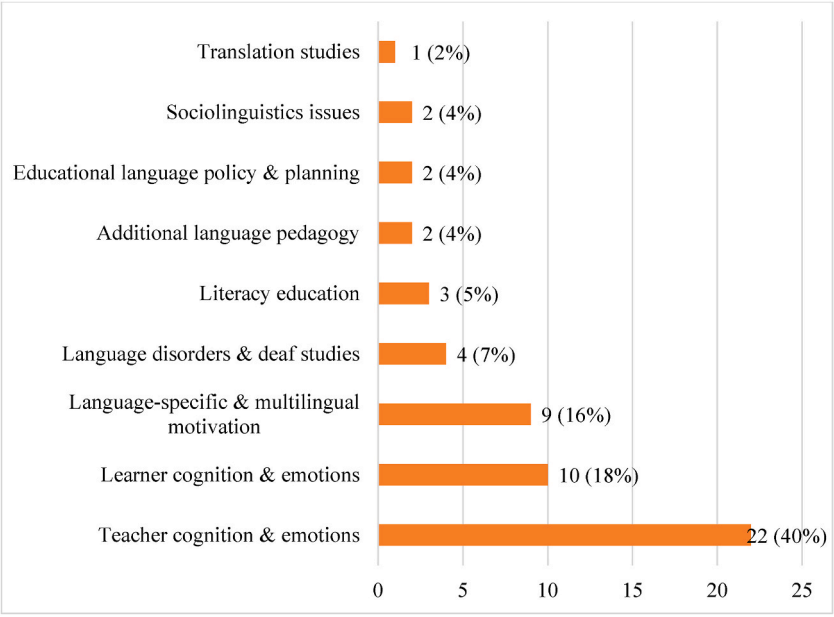


**Fig. 4.** Selected studies by domain.

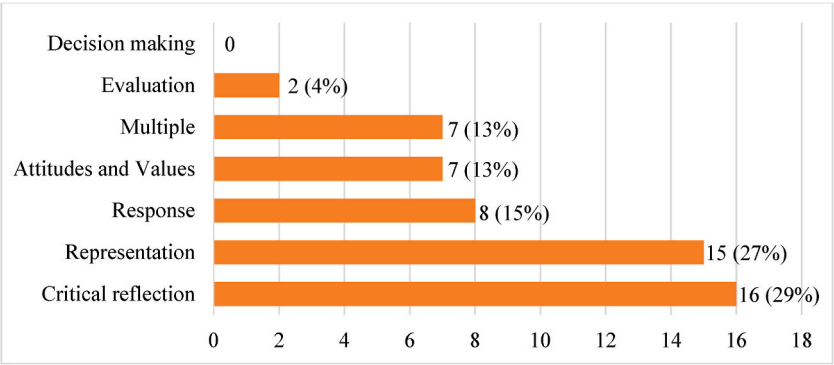


**Fig. 5.** Selected studies by study focus.

attitudes and values (n = 1).

### *3.1.3. Research design*

In terms of research design, we found that 51 studies (93%) adopted a cross-sectional design (i.e., with data generated and analysed at a single time point), followed by merely four studies (7%) designed as longitudinal. Of note is the fact that no study in our sample utilised a single-case design.

## *3.2. RQ2: Q-specific characteristics*

### *3.2.1. Q-sample characteristics*

In this section, we focus on Q-sample development, namely the process of extracting and selecting the Q-items that will form the Q-set, together with the characteristics of the Q-sets used in the selected studies.

*3.2.1.1. Sources of Q-sample development.* As previously discussed, Q-items can be taken from various sources, such as academic literature, newspapers and magazines, or by interviewing members of the target population. AL researchers tended to use multiple sources (n = 37 of the 65 study components examined, or 57%) to construct the Q-set. When a single type of source was used to obtain the Q-set items, this tended to be academic literature (e.g., adapting other previously developed Q-sets or extracting items from other instruments like questionnaires) (n = 19, 29%) and sources classified as *other* (e.g., course tasks, interviews, teachers' reflections, informal talks with students, textbooks and websites). Interestingly, two studies did not provide any information on the sources of their

Q-sets. When multiple sources were used (n = 37), 73% (n = 27) of them comprised of the categories *other* and *academic literature*, whereas in five cases (14%) a combination of academic literature, magazines, policy documents and other sources was used. Less frequently used combinations of sources (n = 5, 14%) were: (a) *other* and *newspapers,* (b) *policy documents* and *other,* (c) *other, policy documents* and *newspapers,* (d) *other, academic literature, newspapers,* and *magazines*, and (e) *academic literature, manuals, magazines* and *other*.

*3.2.1.2. Type of Q-items.* In line with the results obtained in relation to the Q-sample sources, Q-items tended to take multiple forms (here coded as *mixture*) in 27 studies (42%). In 14 of the 27 study components coded as *mixture* (52%), Q-items took the form of either opinions and actions or opinions and strategies, whereas 15% of this subsample used a combination of opinions, actions and experiences. When a single source of Q-items was used, Q-items took the form of opinion statements (n = 20, 31%), followed by actions (n = 5, 8%), strategies (n = 5, 8%), situations (n = 4, 6%) and experiences (n = 3, 5%). One study also did not provide any information regarding the form of Q-items.

*3.2.1.3. Quality assessment.* In the first section of this paper, we mentioned that the selected Q-items can be refined and quality assessed. In our sample, we found only 38 instances (59%) of study components involving some form of quality assessment. Within this subsample, the most frequent procedure for quality-assessing a Q-set was to have it evaluated by a panel of external experts (n = 14, 37%), followed by the use of multiple strategies (n = 12, 32%), piloting (n = 9, 24%) and other approaches (e.g., a translator involved to check the accuracy of the Q-item translation, n = 3, 8%). When multiple strategies of quality assessment used, these tended to involve both external experts and piloting, except for two cases in which a combination of piloting and other approaches was used.

*3.2.1.4. Size of Q-set and item form.* The descriptive statistics relative to the size of Q-sets (i.e., the number of Q-items in each Q-set) are presented in Table 1. The distribution of Q-set sizes was normal, with its skewness being between −2 and +2 (Byrne, 2010; Hair et al., 2010). The Q-set size ranged from 23 to 94, with a mean and standard deviation of 44.94 and 14.51, respectively. Regarding the form in which the Q-set items appeared, written statements were used in most study components (n = 63, 97%), with just one study using visual items in the form of pictures. Finally, one study used a combination of the two (i.e., pictures accompanied by captions).

### *3.2.2. P-sample characteristics*

In this section, we report the characteristics of the P sample (i.e., the participants) in the selected studies. With regard to the sample size, the P-sample size mean was 29.38, with the smallest number of participants being 5 and the largest being 67 (Table 2). The distribution of sample sizes was approximately normal, with skewness between −2 and +2 (Byrne, 2010; Hair et al., 2010).

Regarding the national context in which participants were recruited, out of the 65 study components examined, 13 (20%) were conducted in the United States (US), followed by China (n = 11, 17%), United Kingdom (UK) (n = 9, 14%), and Australia (n = 6, 9%) (Fig. 6). Some studies recruited participants based in various countries, such as Canada, US, Spain, Germany, Australia, France, the UK and Italy (Lu & Geng, 2022), or Hong Kong, Japan, and the UK (Pemberton & Cooker, 2012).

In relation to the study setting, most studies were conducted in educational contexts, with universities and K-12 (i.e., from kindergarten to secondary/high school) being the most common settings (n = 28, 43%, and n = 27, 42%, respectively). Some studies (n = 6, 9%, here coded as *other*) were conducted outside conventional educational contexts, such as deaf communities and teacher associations. Remarkably, in two studies we have not been able to understand the context as no pertaining information was provided.

Regarding the identity of the research participants, in 80% of cases participants were either teachers (n = 26, 40%) or language learners (n = 26, 40%). Other participants included migrant workers, country citizens, professional writers, and psychologists (n = 7, 11%). In some instances (n = 5, 8%), a variety of participants was recruited, such as learners, teachers, parents, and school administrators (Alkhateeb, Al Hamad, & Mustafawi, 2020) and learners, teachers and researchers (Etherington, 2005).

### *3.2.3. Q-sorting characteristics*

*3.2.3.1. Scale range and forced distribution.* In this section we focus on the range of values used in the Q-grids (the grids on which the Q-items are ordered and placed by participants, see Fig. 1). Various scale ranges were used, with −5 to +5 (n = 29, 45%), and - 4 to + 4 (n = 20, 31%) being the most frequently used scale ranges, although ranges such as −6 to +6 (n = 6, 9%) and −3 to +3 (n = 4, 6%) have been occasionally chosen. Of note here is the use of some unusual ranges (n = 3, 5%) like 0 to 10, 1 to 7 or 1 to 5.

Next, we focused on the use of forced distributions. In the first part of this article, we explained that Q researchers tend to opt for a forced distribution, namely the a-priori decision of how many Q-items participants can place under each column of the Q-grid. Our data support this tendency, as a forced distribution was applied in most study components (n = 60, 92%). In four cases, a free distribution was used (i.e., no limit was placed on the number of Q-items that could be placed under each column of the Q-grid), whilst one study used a combination of the two approaches, whereby participants were allowed to place more or fewer items under a column

**Table 1**
Q-set size, descriptive statistics.

| Range | Min | Max | Mean | *SD* | Skewness | *SE* of Skewness |
|---|---|---|---|---|---|---|
| 71 | 23 | 94 | 44.94 | 14.51 | 1.34 | 0.30 |

**Table 2**
P-set size, descriptive statistics.

| Range | Min | Max | Mean | *SD* | Skewness | *SE* of skewness |
|---|---|---|---|---|---|---|
| 62.00 | 5 | 67 | 29.38 | 14.73 | 1.01 | 0.30 |

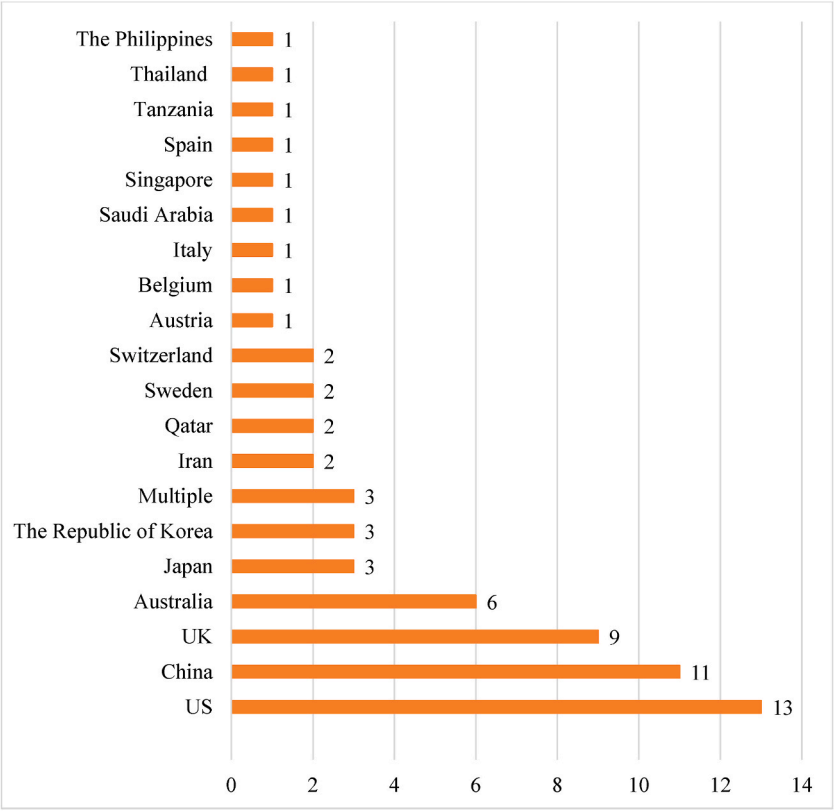


**Fig. 6.** National study context.

if they struggled with a forced distribution.

*3.2.3.2. Modality of sorting activities and post-sorting interviews.* In our sample, most sorting activities were conducted in person (n = 43, 66%), whereas in 15 instances (23%) Q-sorts were generated online. In two cases, a mixture of in-person and online setting was adopted, whereas in three study components (5%) other approaches were used, such as conducting the sorting activity over the phone or via post. When the sorting activities were conducted online, the following programmes and websites were used: a) HtmlQ, b) Q-Sortware, c) Q-TIP, d) QSorTouch, e) Qualtrics, f) Toolset, g) WebQSort, and h) FlashQ application.

Regarding the post-sorting interviews, 41 study components (63%) involved interviewing participants after the sorting activity, whilst in 26% of cases (n = 17) no qualitative data from participants were collected. In some instances in which data collection was not conducted in person (n = 7, 11%), qualitative data were obtained by asking participants to provide written comments about their sorting decisions.

### *3.2.4. Data-analysis characteristics*

*3.2.4.1. Programmes used.* PQMethod (Schmolck, 2014) was the most used piece of software for Q data analysis in our sample (n = 37, 57%), followed by Ken-Q/KADE (n = 12, 19%) (Banasick, 2019). Less frequently used programmes included Stata, PCQ, and QUANL. In six cases (9%), no information was provided.

*3.2.4.2. Factor extraction, rotation, retention, and number of retained factors.* Regarding the factor-extraction methods used, only two techniques were utilised across the 65 study components, namely principal component analysis (n = 22, 34%) and the centroid method (n = 21, 32%). However, it should be highlighted that in a considerable number of cases (n = 22, 34%), no information on factor

extraction was provided.

A similar pattern emerged when examining the factor-rotation techniques that were utilised. Varimax rotation was applied in the majority of the Q factor analyses (n = 45, 69%), which sharply contrasts with the use of judgemental rotation alone, which was employed in only one study component. In six instances (9%) a combination of rotation techniques was used, with all six analyses applying varimax rotation first, followed my judgmental rotation. Finally, once again, in a considerable proportion of study components (n = 13, 20%) the procedures used for factor rotation were not reported.

The factor-retention criteria adopted in the 65 study components are displayed in Fig. 7. In approximately one-third of cases (n = 23, 35%), multiple criteria were used (e.g., eigenvalue>1, number of Q-sorts significantly loading on a factor, interpretability of factors). When a single criterion was followed, the most frequent was retaining any factor with an eigenvalue greater than one (n = 14, 22%). Less frequently chosen approaches for deciding a factor solution were: a) the amount of explained variance (n = 6, 9%), b) the interpretability of factors (n = 5, 8%), c) Humphrey's rule (n = 4, 6%), according to which a factor should be retained "if the cross-product of its two significant loadings (ignoring the sign) exceeds twice the standard error" (Watts & Stenner, 2012, p. 107), and d) the number of Q-sorts significantly loading on a factor (n = 1, 2%). In one case (2%), a factor solution was selected based on whether the solution could distinguish between groups of respondents (Paradice, 2001). Finally, no information was provided in 11 study components (17%).

The number of retained factors ranged from 1 to 7 ($M = 3.42$, $SD = 1.20$) (see Table 3). As displayed in Table 4, 85% of the 65 Q factor analyses conducted across our sample of 55 studies yielded between two and four factors (two factors: n = 13, 20%; three factors: n = 24, 37%; four factors: n = 18, 28%).

*3.2.4.3. Flagging criteria and treatment of confounded Q-sorts.* *Flagging* refers to the action of selecting which Q-sorts will be used to calculate the factor scores of the retained factors. In our sample, the criteria used for flagging Q-sorts were reported in 55% of study components (n = 36), and not provided in the remaining 29 (45%). A related consideration when flagging Q-sorts is how to treat any confounded Q-sorts, namely Q-sorts with a statistically significant factor loading on more than one factor. In 37% of study components (n = 24), no information was provided in this regard. In the remaining components (n = 41), the most common practice was to systematically exclude any confounded Q-sorts (n = 20, 49%), followed by the opposite approach, that is, to include confounded Q-sorts in the analysis by ignoring their secondary factor loading(s) (n = 12, 29%). In nine instances (22%), other approaches were used, the most frequent of which was raising the threshold of significance (so that the secondary significant factor loading of the confounded Q-sorts would fall below the new threshold).

*3.2.4.4. Variance explained.* In 56 of the 65 study components (86%), the authors reported information regarding the amount of variance explained by the factor solution. The mean explained variance was 54% (SD = 10.08; min = 35; max = 79), with a skewness of 0.71 and a standard error of 0.32, indicating a normal distribution (Fig. 8).

*3.2.4.5. Factor interpretation.* The last variable we considered was whether the retained factors were interpreted and described in each study component. Whilst a qualitative interpretation of the factors was given in most cases, in seven instances (11% of the study components) either no factor interpretation was provided or this was deemed inadequate (e.g., only one factor was interpreted).

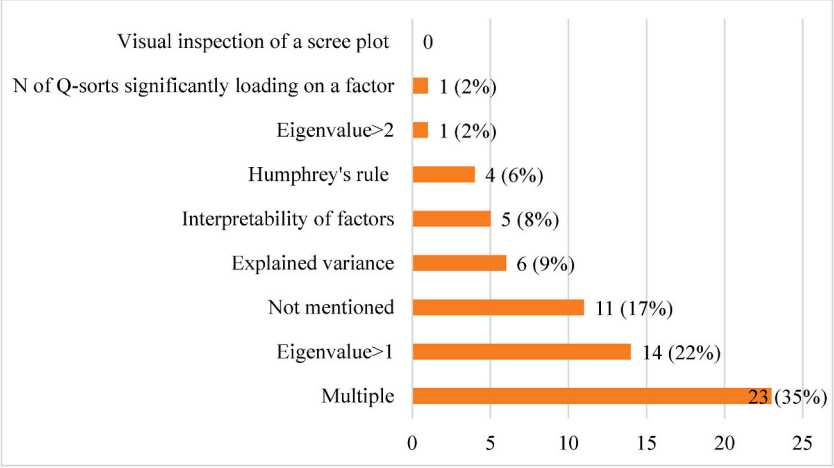


**Fig. 7.** Breakdown of factor-retention approaches used.

**Table 3**
Number of retained factors, descriptive statistics.

| Range | Min | Max | Mean | *SD* | Skewness | SE of skewness |
|---|---|---|---|---|---|---|
| 6 | 1 | 7 | 3.42 | 1.20 | .77 | .30 |

**Table 4**
Breakdown of number of retained factors.

| Number of retained factors | Number of study components (n = 65) | % |
|---|---|---|
| 1 | 1 | 2 |
| 2 | 13 | 20 |
| 3 | 24 | 37 |
| 4 | 18 | 28 |
| 5 | 4 | 6 |
| 6 | 4 | 6 |
| 7 | 1 | 2 |

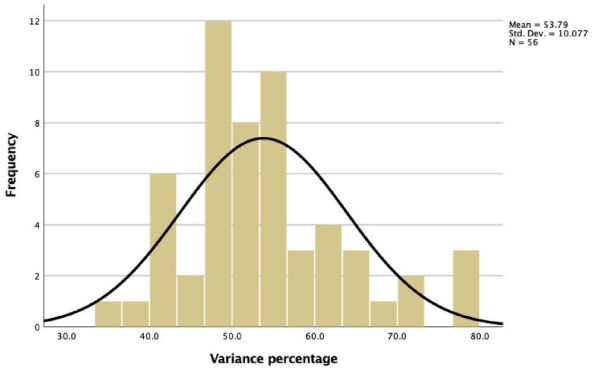


**Fig. 8.** Histogram of explained variance.

## 4. Discussion and conclusion

### *4.1. A rapidly growing methodology in applied linguistics*

This article was aimed at providing a comprehensive account of the applications of Q methodology in AL research. This was accomplished by reviewing a sample of 55 empirical studies that used Q to investigate a language-related issue or phenomenon, and by providing an overview of their contextual and methodological characteristics.

This review revealed that Q is growing in popularity as a methodology to investigate people's subjectivity around language-related issues or phenomena. This was evident by the rapid increase in Q studies in the field since 1999, with more Q studies having been published in the last four years (2019–July 2022, n = 32) than in the previous two decades (1999–2018, n = 23). The fact that 16% of studies in our sample were graduate theses and dissertations also indicates that Q is becoming increasingly used by postgraduate students and known within academic research programmes. Overall, this is in line with the findings from systematic reviews in other research fields. For example, in a review of Q research in compulsory education in the decade 2010–2019, Lundberg et al. (2020) found that 35 of the 74 selected studies were published between 2017 and 2019. Similarly, 148 of the 289 studies reviewed by Churruca et al. (2021) in the field of healthcare research (whose publishing year ranged from 1966 to 2019) were published between 2015 and 2019.

This review also showed that most empirical Q studies in AL were set in an educational context. Most commonly, Q has been used to investigate teachers' and learners' cognition and emotions in relation to a language-related issue or phenomenon. Q has also been frequently used to explore language learners' and teachers' motivation. Outside of education, Q has been employed to conduct research related to sociolinguistics issues, translation studies, language disorders and deaf studies. This indicates that Q is gaining popularity in various areas within the field, but that it has most prominently been exploited in language-education research. When comparing these areas of language-related research with the domains suggested by Irie et al.(2018) (i.e., affects, beliefs, identity and agency) and Thumvichit (2022a) (i.e., choice, emotions, motivations and values), this review showed that while studies focusing on emotions, beliefs and motivations were common, limited research using Q has been conducted in the areas of identity and agency, both from the perspective of language learners and teachers. In this regard, we acknowledge that the work by Wu and Forbes (2023, 2022), published after our database search, has investigated the construct of language learners' multilingual identity using Q (see below).

In terms of study focus, the reviewed studies tended to adopt Q as a tool to allow research participants to critically reflect on

themselves as language teachers or learners (e.g., on their motivation for language teaching/learning, on their strategies for self-directed learning, on their emotions related to language teaching/learning, on their perceptions of their language skills). At the same time, an almost equally widespread focus was participants' opinions regarding a language-related issue (e.g., stakeholders' perspectives on educational language policies, teachers' beliefs about multilingualism and linguistically inclusive pedagogies). Whilst these findings broadly resonate with Lundberg et al. (2020), who found that representation, attitudes and values, and critical reflection were the most widespread study foci in their sample of 74 studies, it seems that, in AL, Q has been considered particularly useful for encouraging participants to critically reflect on their own situation, experience and characteristics. This finding echoes Irie et al.'s (2018) and Pemberton and Cooker's (2012) argument that the sorting activity is a particularly suitable tool for fostering participant reflexivity and self-awareness. In this regard, we encourage researchers conducting intervention studies to consider using the Q-sort method not only as an instrument of data-collection and analysis, but also as an integrative part of the intervention being tested.

We acknowledge that a limitation of this research synthesis is the inclusion of studies published and indexed up to July 2022 (when the database search was conducted). To reduce this limitation, a database search was conducted in October 2023 to identify relevant research published between August 2022 and October 2023. The literature search led to the identification of 14 empirical studies, which were examined in light of the findings from this review. Firstly, the number of studies identified (n = 5 between August–December 2022 and n = 9 between January–October 2023) confirms the rapid growth of Q research in AL. In terms of study domain, the pattern emerging from these studies closely aligns with our findings, with teacher cognition and emotions representing the most common research area, followed by learner cognition and emotions, and language-specific and multilingual motivation (Fig. 9). The widespread focus on teachers' and, to a lesser extent, learners' emotions (e.g., Ding, Liu, & Peng, 2023; Fraschini, 2023; Thumvichit, 2022b, 2023a, 2023b), and on language learners' motivation and identity (i.e., Lu, Wang, Shen, & Gao, 2022; Wu & Forbes, 2022, 2023) confirms that Q continues to be primarily used in the field as a tool to both allow participants to critically reflect on their personal experiences and to explore participants' attitudes and values.

Based on the findings from this study, and in light of relevant research published after this review, we now provide directions for future research.

### 4.2. *Directions for future research*

This review has found some gaps in the current applications of Q methodology, which reveal possible directions for future research. Firstly, we found that a relatively high proportion of studies omitted important information in relation to the analytical decisions used to analyse Q-sort data. In a noticeable proportion of study components, the authors failed to report information related to the Q factor-analysis procedures, such as the criteria used for factor extraction (34%), retention (17%) and rotation (20%), together with the criteria used for selecting Q-sorts for factor-score calculation (45%) and the treatment of confounded Q-sorts (37%). Furthermore, important stages of Q were occasionally ignored, namely conducting post-sorting interviews with research participants (26%) and even qualitatively interpreting each retained factor (11%). We hypothesise that a reason for some of these omissions may be a reluctance to include technical information related to Q factor analysis. Considering the likely unfamiliarity of the readership of these studies with Q methodology and its analytical procedures, the authors and/or editors may have preferred to avoid discussing technical issues such as Q-sort flagging and treatment of confounded Q-sorts. Nonetheless, reporting such information is essential not only to ensure methodological transparency and enable replicability of Q studies, but to also increase the quality and methodological rigour of published Q research. In consideration of the number of qualitative decisions involved in the process of Q factor analysis, we thus recommend both authors and reviewers to provide and expect a discussion and justification of the chosen analytical approaches.

As previously discussed, the lack of shared practices during the process of Q-set development has been considered a potential source of bias in Q research (Kampen & Tamás, 2014). In this regard, our analysis showed that a noticeable proportion of the analysed studies did not involve or mention any form of quality assessment during the stage of Q-set development. Considering the complex and iterative nature of developing a new Q-set, it is recommended that multiple forms of quality assessment are adopted. We particularly encourage the involvement of a panel of experts, both Q methodologists and experts in the topic of inquiry, to assess the

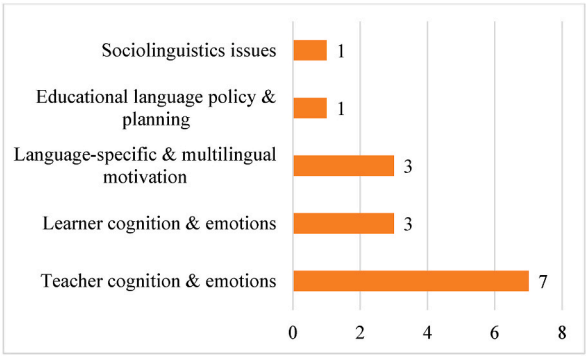


**Fig. 9.** Domain of identified studies published between August 2022 and October 2023.
*Note.* One paper (Lu & Xiong, 2023) focuses on both students and teachers and was thus counted in both domains.

appropriateness and comprehensiveness of the Q-set, together with a piloting phase actively involving participants not only in the sorting activity, but also in providing feedback on the extent to which the Q-set enabled them to fully express their viewpoints. Furthermore, researchers should consider publishing methodological papers focusing on the process of Q-set development, which are currently lacking in the field, as this would provide examples of good practice. In this regard, we found only one instance of published research focusing on the process of Q-sort development (Wang, Nikitina, Kaur, & Furuoka, 2022).

From a research-design perspective, most of the reviewed studies have used Q to explore the subjectivity of a group of participants at a single time point. As demonstrated in Morea (2022), Q can also be adopted in research designs involving a temporal variable, such as longitudinal and experimental research. Q can thus be effectively used to both track the development of participants' viewpoints over time and assess the effectiveness of interventions. Additionally, Q has been traditionally conceived as a methodology for the single case (Stephenson, 1953), and applications of Q in single-case designs are not unusual in other research fields. However, none of the studies in our sample have employed such a research design, a finding also in line with Lundberg et al. (2020). We acknowledge that Fraschini (2022) has recently conducted a longitudinal, single-case study on learners' emotional dynamics in instructed language learning contexts, yet this seems to represent the only occurrence in the field. Single-case Q research designs represent a promising approach for tracking individuals' trajectories of belief, emotion or identity change and stability over a specific period of time (e.g., a language course, a study-abroad experience, a teacher-education programme).

Finally, another finding with potential for informing future practice is the limited range of factor extraction and rotation techniques employed by AL researchers. Regarding factor extraction, only principal component analysis and the centroid method were used in the reviewed studies. Whilst these seem to be the most common factor-extraction methods used in other research fields (Lundberg et al., 2020; Morea, 2022), other techniques have been suggested, such as principal axis factoring and maximum likelihood factor extraction (Akhtar-Danesh, 2017), with the former being a particularly versatile technique when extracting factors for theoretical purposes (Akhtar-Danesh, 2017). It can be hypothesised that the limited range of factor-extraction methods used may be a reflection of the options available in the most utilised programmes for Q factor analysis. Regarding factor rotation, varimax was used in over two-thirds of Q factor analyses in our sample. In line with Akhtar-Danesh's (2016) suggestion, we recommend researchers to consider other rotation techniques like quartimax, equamax, direct oblimin and promax, with each having its own features. In particular, quartimax may be particularly suited to Q: since it "minimises the number of factors to explain each variable" (i.e., each Q-sort) (Akthar-Danesh, 2016, p. 6), this technique may reduce the number of confounded Q-sorts, as each Q-sort will be "loaded on the minimum number of factors" (p. 6).

### 4.3. Limitations

Although the present review was conducted following rigorous procedures for identifying and selecting studies, it is expected that not all Q studies in the field of AL have been retrieved. Firstly, potentially eligible studies may have not been indexed in the databases inspected (particularly dissertations and theses). Secondly, only studies in the English language were included, resulting in an over-representation of studies conducted in anglophone countries. Additionally, only studies indexed on the reviewed databases up to July 2022 were considered, thus excluding any potentially eligible study published more recently.

### 4.4. Conclusion

Despite an increase in conceptual papers arguing for the potential of Q methodology for AL, and a growing number of empirical Q studies in language-related research, we found a gap in systematic research syntheses mapping the applications of Q in the field. As a result, this review may be useful not only to applied linguists unfamiliar with Q or who may be considering using this methodology, but also to researchers currently employing Q and who may be interested in comparing their study design and analytical decisions with other research in the field.

Although this review has shown that Q methodology is growing in popularity in applied linguistics, its applications are still relatively limited compared to other methods like questionnaires and interviews. By mapping the applications of Q within the field to this day, it is our hope that this review will draw applied linguists' attention to an almost century-old, but still relatively novel methodology that can open new possibilities for systematically exploring people's subjectivity around language-related issues and phenomena.

## CRediT authorship contribution statement

**Nicola Morea:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hessameddin Ghanbar:** Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

## Declaration of competing interest

The authors declare none.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.system.2023.103194.

## References


Akhtar-Danesh, N. (2017). A comparison between major factor extraction and factor rotation techniques in Q-methodology. *Open Journal of Applied Sciences, 7*, 147–156. https://doi.org/10.4236/ojapps.2017.74013

Akthar-Danesh, N. (2016). An overview of the statistical techniques in Q methodology: Is there a better way of doing Q analysis? *Operant Subjectivity: The International Journal of Q Methodology, 38*(3–4), 1–8.

Alanazi, A., Wharrad, H., Moffatt, F., Taylor, M., & Ladan, M. (2021). Q methodology in the COVID-19 era. *Healthcare, 9*(11). https://doi.org/10.3390/healthcare9111491

* Alkhateeb, H., Al Hamad, M., & Mustafawi, E. (2020). Revealing stakeholders' perspectives on educational language policy in higher education through Q-methodology. *Current Issues in Language Planning, 21*(4), 415–433. https://doi.org/10.1080/14664208.2020.1741237.

Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing, 38*(1), 6–40. https://doi.org/10.1177/0265532220927487

Banasick, S. (2019). Kade: A desktop application for Q methodology. *Journal of Open Source Software, 4*, 1–4. https://doi.org/10.21105/joss.01360

* Boscolo, P., & Cisotto, L. (1999). Instructional strategies for teaching to write: A Q-sort analysis. *Learning and Instruction, 9*(2), 209–221. https://doi.org/10.1016/S0959-4752(98)00042-5.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Brown, S. R. (1993). A *primer* on *Q methodology*. *Operant Subjectivity, 16*, 91–138.

Brown, S. R. (2006). A match made in heaven: A marginalized methodology for studying the marginalized. *Quality and Quantity, 40*, 361–382. https://doi.org/10.1007/s11135-005-8828-2

Bryfonski, L., & McKay, T. H. (2019). TBLT implementation and evaluation: A meta-analysis. *Language Teaching Research, 23*(5), 603–632. https://doi.org/10.1177/1362168817744389

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Routledge.

Chong, S., & Plonsky, L. (2023). A typology of secondary research in Applied Linguistics. *Applied Linguistics Review*, 1–26. https://doi.org/10.1515/applirev-2022-0189

Churruca, K., Ludlow, K., Wu, W., Gibbons, K., Nguyen, H. M., Ellis, L. A., et al. (2021). A scoping review of Q-methodology in healthcare research. *BMC Medical Research Methodology, 21*(1). https://doi.org/10.1186/s12874-021-01309-7

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approach* (2nd ed.). Sage Publications.

Crowther, D., Kim, S., Lee, J., Lim, J., & Loewen, S. (2021). Methodological synthesis of cluster analysis in second language research. *Language Learning, 71*(1), 99–130.

Dieteren, C. M., Patty, N. J. S., Reckers-Droog, V. T., & van Exel, J. (2023). Methodological choices in applications of Q methodology: A systematic literature review. *Social Sciences & Humanities Open, 7*(1). https://doi.org/10.1016/j.ssaho.2023.100404

Ding, X., Liu, Y., & Peng, J. (2023). Straddling burnout and resilience: A Q methodology study among high school EFL teachers. *Journal of Multilingual and Multicultural Development*. https://doi.org/10.1080/01434632.2023.2193171

* Etherington, M. S. (2005). *Student perceptions of grammar and its role within the UK university English for Academic Purposes context [Doctoral dissertation, University of Manchester]*. ProQuest Dissertations & Theses Global.

Fraschini, N. (2023). Language learners' emotional dynamics: Insights from a Q methodology intensive single-case study. Language, Culture and Curriculum, 2. In J. Hair, W. C. Black, B. J. Babin, & R. E. Anderson (Eds.), *Multivariate data analysis* (7th ed., vol. 36, pp. 222–239). Pearson Educational International, 2010.

Ghanbar, H., & Rezvani, R. (2023). Structural equation modeling in L2 research: A systematic review. *International Journal of Language Testing, 13*(Special Issue), 79–108. https://doi.org/10.22034/IJLT.2023.381619.1224

Hensel, D., Toronto, C., Lawless, J., & Burgess, J. (2022). A scoping review of Q methodology nursing education studies. *Nurse Education Today, 109*. https://doi.org/10.1016/j.nedt.2021.105220

In' nami, Y., & Koizumi, R. (2011). Structural equation modeling in language testing and learning research: A review. *Language Assessment Quarterly, 8*(3), 250–276.

* Irie, K., Ryan, S., & Mercer, S. (2018). Using Q methodology to investigate pre-service EFL teachers' mindsets about teaching competences. *Studies in Second Language Learning and Teaching, 8*(3), 575–598.

Kampen, J. K., & Tamás, P. (2014). Overly ambitious: Contributions and current status of Q Methodology. *Quality and Quantity: International Journal of Methodology, 48*(6), 3109–3126. https://doi.org/10.1007/s11135-013-9944-z

Kostromitina, M., & Plonsky, L. (2022). Elicited imitation tasks as a measure of L2 proficiency: A meta-analysis. *Studies in Second Language Acquisition, 44*(3), 886–911. https://doi.org/10.1017/S0272263121000395

Li, X. (2022). The significance of Q-methodology as an innovative method for the investigation of affective variables in second language acquisition. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.995660

* Lu, X., & Geng, Z. (2022). Faith or path? Profiling the motivations of multilingual Chinese as a foreign language teachers using Q methodology. *Language Teaching Research, 0*(0). https://doi.org/10.1177/13621688221099079.

Lundberg, A., de Leeuw, R., & Aliani, R. (2020). Using Q methodology: Sorting out subjectivity in educational research. *Educational Research Review, 31*. https://doi.org/10.1016/j.edurev.2020.100361

Lu, J., Wang, Y., Shen, Q., & Gao, X. (2022). Investment in learning Chinese by international students studying Chinese as a second language (CSL). *Sustainability, 14*. https://doi.org/10.3390/su142315664

Lu, J., & Xiong, T. (2023). The attitudes of high school students and teachers toward mobile apps for learning English: A Q methodology study. *Social Sciences & Humanities Open, 8*(1). https://doi.org/10.1016/j.ssaho.2023.100555

Marsden, E., Mackey, A., & Plonsky, L. (2016). The IRIS repository: Advancing research practice and methodology. In A. Mackey, & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS repository of instruments for research into second languages* (pp. 1–21). Taylor and Francis. https://doi.org/10.4324/9780203489666.

Meehan, K., Ginart, L., & Ormerod, K. J. (2022). Sorting at a distance: Q methodology online. *Field Methods, 34*(1), 82–88. https://doi.org/10.1177/1525822X211069657

* Morea, N. (2022). Investigating change in subjectivity: The analysis of Q-sorts in longitudinal research. *Research Methods in Applied Linguistics, 1*(3). https://doi.org/10.1016/j.rmal.2022.100025.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 2021*, 372. https://doi.org/10.1136/bmj.n71

* Paradice, R. (2001). An investigation into the social construction of dyslexia. *Educational Psychology in Practice, 17*(3), 213–225. https://doi.org/10.1080/02667360120072747.

Park, H. I., Solon, M., Dehghan-Chaleshtori, M., & Ghanbar, H. (2022). Proficiency reporting practices in research on second language acquisition: Have we made any progress? *Language Learning, 72*(1), 198–236. https://doi.org/10.1111/lang.12475

* Pemberton, R., & Cooker, L. (2012). Self-directed learning: Concepts, practice, and a novel research methodology. In S. Mercer, S. Ryan, & M. Williams (Eds.), *Psychology for language learning: Insights from research, theory, and practice* (pp. 203–219). Palgrave Macmillan.
Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition, 35*, 655–687. https://doi.org/10.1017/S0272263113000399
Plonsky, L. (2014). Study quality in quantitative L2 research (1990–2010): A methodological synthesis and call for reform. *The Modern Language Journal, 98*, 450–470. https://doi.org/10.1111/j.1540-4781.2014.12058.x
Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. *The Modern Language Journal, 102*(4), 713–731. https://doi.org/10.1111/modl.12509
Plonsky, L., & Gonulal, T. (2015). Methodological synthesis in quantitative L2 research: A review of reviews and a case study of exploratory factor analysis. *Language Learning, 65*(S1), 9–36.
Plonsky, L., & Kim, Y. (2016). Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics, 36*, 73–97. https://doi.org/10.1017/S0267190516000015
Plonsky, L., Marsden, E., Crowther, D., Gass, S. M., & Spinner, P. (2020). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research, 36*(4), 583–621. https://doi.org/10.1177/0267658319828413
Ramlo, S. (2016). Mixed method lessons learned from 80 Years of Q methodology. *Journal of Mixed Methods Research, 10*(1), 28–45. https://doi.org/10.1177/1558689815610998
Riazi, M., Shi, L., & Haggerty, J. (2018). Analysis of the empirical research in the journal of second language writing at its 25th year (1992–2016). *Journal of Second Language Writing, 41*, 41–54. https://doi.org/10.1016/j.jslw.2018.07.002
Schmolck, P. (2014). *PQ Method manual*. http://schmolck.userweb.mwn.de/qmethod/pqmanual.htm.
Stephenson, W. (1953). *The study of behavior; Q-technique and its methodology*. University of Chicago Press.
Sudina, E. (2021). Study and scale quality in second language survey research, 2009–2019: The case of anxiety and motivation. *Language Learning, 71*, 1149–1193. https://doi.org/10.1111/lang.12468
Sudina, E. (2023). *Scale quality in second-language anxiety and WTC: A methodological synthesis*. https://doi.org/10.1017/S0272263122000560. *Studies in Second Language Acquisition*, 1–29.
Thumvichit, A. (2022a). Illuminating insights into subjectivity: Q as a methodology in applied linguistics research. *Applied Linguistics Review*. https://doi.org/10.1515/applirev-2021-0205
Thumvichit, A. (2022b). Enjoyment in language teaching: A study into EFL teachers' subjectivities. *International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1515/iral-2022-0087
Thumvichit, A. (2023a). 'I'm aware of that, but…': Breaking the silence on moral distress among language teachers. *Language Culture and Curriculum, 36*(3), 343–360. https://doi.org/10.1080/07908318.2023.2189268
Thumvichit, A. (2023b). Understanding emotion-regulation strategies among foreign language teachers in response to classroom stressors: Findings from a Q methodology study. *Social Psychology of Education*. https://doi.org/10.1007/s11218-023-09809-8
Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: "Clozing" the gap. *Studies in Second Language Acquisition, 33*, 339–372. https://doi.org/10.1017/S0272263111000015
Vuogan, A., & Li, S. (2023). A systematic review of meta-analyses in second language research: Current practices, issues, and recommendations. *Applied Linguistics Review*. https://doi.org/10.1515/applirev-2022-0192
Wang, Y., Nikitina, L., Kaur, J., & Furuoka, F. (2022). Developing and piloting a Q-sample on Chinese language learners' epistemic beliefs. *Journal of Language and Education, 8*(3), 119–132. https://jle.hse.ru/article/view/13590.
Watts, S., & Stenner, P. (2012). *Doing Q methodological research: Theory, method and interpretation*. SAGE Publications Ltd.
Wu, X., & Forbes, K. (2022). Multilingual identity in schools: A Q methodological study with Chinese LOTE-as-L3 learners in international and public high school contexts. *Journal of Language, Identity and Education*. https://doi.org/10.1080/15348458.2022.2134138
Wu, X., & Forbes, K. (2023). Contextualising the temporal dimension in multilingual identity construction: A longitudinal Q methodological study with Chinese LOTE-learning public high school students. *System, 113*. https://doi.org/10.1016/j.system.2023.103090
Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing, 33*, 497–528.
Zhang, M., Gibbons, J., & Li, M. (2021). Computer-mediated collaborative writing in L2 classrooms: A systematic review. *Journal of Second Language Writing, 54*. https://doi.org/10.1016/j.jslw.2021.100854
Zhang, M., & Plonsky, L. (2020). Collaborative writing in face-to-face settings: A substantive and methodological review. *Journal of Second Language Writing, 49*. https://doi.org/10.1016/j.jslw.2020.100753
included in the review