

*Investigating individual differences
with qualitative research methods: Results of
a meta-analysis of leading applied linguistics journals*

Ágnes Albert

Eötvös Loránd University, Budapest, Hungary

<https://orcid.org/0000-0001-8339-7119>

albert.agnes@btk.elte.hu

Kata Csizér

Eötvös Loránd University, Budapest, Hungary

<https://orcid.org/0000-0003-1755-8142>

wein.kata@btk.elte.hu

Abstract

The aim of the present article is to provide a systematic review of qualitative studies in the leading journals of our field focusing on their distributional properties in the various journals as well as topic choice and selected quality control issues. In order to achieve this aim, we carried out a systematic review of research articles published in leading journals in our field, namely, *Applied Linguistics*, *Language Learning*, *Language Teaching Research*, *Studies in Second Language Acquisition* and *Modern Language Journal* between 2016 and 2020. Our sample contains 93 articles in which researchers employed qualitative research methods or mixed methods including a qualitative component. Our main results indicate that there is great variation among journals in terms of the number of qualitative studies. As for topic considerations, some traditional individual difference variables seem to have a dominant role, with cognitive processes involved in language acquisition gaining some ground as well. Concerning quality control issues, there could be room for improvement with regard to reporting the quality control measures, including the tools employed in the studies. Based on our results, we can conclude that a more systematic understanding of acceptable processes in

the field of applied linguistics could increase not only the number of qualitative articles published but also their topical importance.

Keywords: qualitative research; individual differences; quality control; meta-analysis; systematic review

1. Introduction

Applied linguistics research has long been centered on individual differences (IDs) among learners and why and how these differences influence the learning processes and outcomes. Comparing the two editions of Zoltán Dörnyei's comprehensive monograph on individual difference research (Dörnyei, 2005; Dörnyei & Ryan, 2015), it becomes apparent that the qualitative approach, which "uses text as empirical material (instead of numbers), starts from the notion of the social construction of realities under study, and is interested in the perspectives of participants, in everyday practices and everyday knowledge referring to the issue under study" (Flick, 2018, p. 2), is gaining ground in the investigation of these variables. This is because researchers are moving away from large-scale data collection to more situated and contextualized studies that are focusing on actual differences of learners instead of providing often vaguely generalizable results that might not be relevant in many contexts and learning environments. Despite this seeming interest in qualitative methods, there are only few comprehensive reviews of qualitative studies (see Chong & Plonsky, 2021a, 2021b), while both content and quality-related meta-analyses of quantitative research appear more often in leading journals (e.g., Plonsky & Derrick, 2016). Hence, the aim of this study is to provide a systematic review of qualitative studies in the leading journals of our field, that is, *Applied Linguistics*, *Language Learning*, *Language Teaching Research*, *Studies in Second Language Acquisition*, and *Modern Language Journal*. The importance of our research lies in its potential to inform future studies concerning possible quality control measures to be applied in qualitative investigations. In this article, we focus on multiple issues starting with the descriptive analysis of the distribution of the papers in these journals and the topics they cover. Next, we analyze what type of data collection methods and techniques researchers used. Finally, we touch upon quality control issues in connection with three of the data collection tools.

2. Literature review

2.1. ID variables and their investigation

The investigation of IDs is a diverse field including an ever-increasing number of variables that researchers think relevant for students and teachers in various

contexts. Some of the classic constructs include language learning aptitude that explains the rate of acquisition of language learners (Carroll & Sapon, 1959), language learning motivation that measures the amount of effort students are willing to invest in language learning (Dörnyei & Ushioda, 2011) or learners' age and its influence on learning processes and outcomes (e.g., Pfenninger & Singleton, 2021). In addition, language learning styles and strategies have traditionally been seen important in the learning process, the former indicating the preferred cognitive styles used by students when processing information and completing tasks (Reid, 1995), while the latter includes those conscious techniques that students might apply when learning or solving a task (O'Malley & Chamot, 1990). A further cognitive variable are language learning beliefs that concern those dispositions that students or teachers hold in connection with the target language, its learning, or themselves (Dörnyei, 2005; Mori, 1999). It has also been acknowledged that emotions play an important role in shaping learning. One influential emotion, anxiety, has widely been researched (Horwitz et al., 1986), while the inclusion of further negative (e.g., shame, Teimouri, 2017) and positive (e.g., enjoyment, Dewaele & MacIntyre, 2016) emotions is a more recent advancement in our field. It has also been acknowledged that successful language learning is not possible without students taking responsibility for their own learning and regulating this process (Kormos & Csizér, 2014), which needs to be coupled with their willingness to communicate (MacIntyre et al., 1998). In terms of the most recent addition to the list of ID variables, they include experience-related notions, such as the flow experience, that is, immersing oneself into the learning and forgetting about the passing of time (Csikszentmihályi et al., 2005), or engagement, which is a complex construct, including a range of psychological events, a certain quality of interaction, and positive emotions (Shernoff, 2013). Although ID variables are mostly investigated individually, Ryan's (2019) recent call for the need to investigate such factors in concert in order to understand their interplay and the way they shape one another in complex ways should definitely be considered.

It is an undeniable fact that the investigation of ID variables has traditionally employed quantitative questionnaire studies that provide cross-sectional analyses of individuals, whose answers could be generalized to a wider population. While questionnaires are versatile and economical instruments to collect data, there are a number of possible critical considerations that show the limitations of such studies, which led to the use of qualitative investigations within the field of ID research (Dörnyei & Ryan, 2015). However, given the diversity of these variables and the fact that researchers usually tend to concentrate on a limited number of them, research strategies might differ for different ID variables. Examples of addressing such limitations come from subfields in which qualitative

techniques have been employed to counter-balance certain shortcomings within the field. For instance, in the L2 motivation field, Ushioda (2001) started to employ qualitative methods to map the longitudinal changes in L2 motivation and its relation to language learning autonomy. The number of qualitative studies has steadily increased in this field, which has resulted in a more balanced approach to L2 motivational issues (Dörnyei & Ryan, 2015). A similar tendency can be observed in the investigation of language learning strategies, as Woodrow (2005) called our attention to the fact that in order to investigate learning strategies in a situated manner, more qualitative studies are needed. In fact, the special issue of *System* published in 2014 already contains quite a few qualitative studies. Another such example is provided by mapping learners' willingness to communicate (WTC). First, quantitative studies were used to operationalize and investigate this concept, but these later gave way to innovative qualitative studies in order to research the personal differences in WTC and its subtle temporal changes (MacIntyre et al., 2011; MacIntyre & Legatto, 2011).

Another important issue that needs to be considered when looking at the methodological issues pertaining to the investigation of ID variables is the extent to which a given variable is integrated into the field. The problem is the fact that although second language acquisition theories have long included a number of cognitive variables, such as attention, noticing or awareness, in explaining learning processes, the empirical investigation of such variables as ID concepts has not drawn much attention (Robinson, 1995, 2012; Schmidt, 1990, 2010). The process-like characteristics of these variables would call for longitudinal and exploratory qualitative studies, which are not unheard of in our field but their meaningful combination is still a task ahead of us.

Another issue that needs to be mentioned here is the investigation of emotions. Despite the fact that anxiety and its role in different aspects of learning has been researched in much detail (Horwitz et al. 1986; MacIntyre & Gardner, 1994), the efforts to include additional negative and positive emotions are recent. Csizér, Albert and Piniel (2021; see also Albert & Piniel, 2021; Albert et al., in press) investigated the role of various negative and positive emotions such as, for example, pride, hope, curiosity, shame, boredom, and apathy. One of their most important results concerns the fact that the supreme role of anxiety cannot be proved in all contexts, and other negative emotions play equally strong roles. In addition, in comparative studies of positive and negative emotions, the role of the former seems to be more defining.

The next issue pertaining to ID research that should be considered are the trait- and state-characteristics of ID factors. In the traditional operationalization of ID variables, there seems to be agreement that these are trait-like characteristics which are individually consistent over time (Dörnyei, 2005). However, drawing

on Long (2014), for example, one cannot neglect the fact that many ID variables behave in a state-like manner and are susceptible to change due to contextual influences. In one of our recent theoretical overview papers (Csizér & Albert, 2021), it was shown that there are still only few investigations that focus on state-like characteristics of ID variables. As we argued, this could be best achieved in complex qualitative and mixed methods studies, by including multiple sources of data within the same project.

Despite the fact the aim of any large-scale investigation is to obtain results that would be generalizable to the population from which participants were selected, Lowie and Verspoor (2019) convincingly show us that students scoring similarly on ID scales can still be strikingly different when the actual learning processes are also considered. In this sense, naming the field *individual differences* is really a misnomer as researchers up until very recently were more interested in the ways learners are similar to one another (Dörnyei, 2005; Dörnyei & Ryan, 2015). Therefore, when investigating ID variables, specific learning processes should also be included in empirical research, and qualitative designs lend themselves well to this purpose.

These considerations all point to the conclusion that qualitative investigations should be fully integrated into the investigation of ID variables, but this cannot be done without considering what quality control issues should be taken into consideration. As multiple issues should be tackled when designing and executing such studies, the next section provides an overview of quality control in qualitative studies.

2.2. Quality control in qualitative research

Assuring that the quality of qualitative research meets given standards so that the findings of the study can be trusted is an important issue, discussed in most textbooks dealing with research methods (see e.g. Creswell, 2002, 2018; Delamont, 2012; Dörnyei, 2007; Fraenkel et al., 2012; Heigham & Croker, 2009; Patton, 2015; Tracy, 2019). Nevertheless, establishing precisely *what* the basic ideals are that good qualitative research should adhere to and *how* those can be met in terms of specific actions are much less clear. The conceptual framework and the means to ensure good quality research seem to be more straightforward in quantitative studies: the overarching terms of validity and reliability and their related concepts and subcategories comprise the theoretical framework while different sampling and statistical procedures provide the means for implementing those ideals (Dörnyei, 2007). Although, according to Mirhosseini (2020), students or beginner qualitative researchers who were mainly trained in the positivist tradition might believe that the ideals of objectivity, reliability and generalizability can be extended to and applied in qualitative research, he claims that this approach goes against the very nature of qualitative enquiry as “qualitative research is not

unbiased, replicable, and generalizable" (p. 178). Hence, in what follows, we provide a succinct overview of the most important, overall quality control issues, but, as in this study we mainly concentrate on the most often used data collection tools, the section ends with a discussion of tool-related quality insurance.

The initial approach to quality control in qualitative studies involves researchers trying to adhere to essentially quantitative principles of quality. This approach is apparent in Guba's (1981) famous work, in which he reinterpreted and relabeled several cornerstones of positivist ideals of good quality science. In order to ensure the trustworthiness of qualitative research, he came up with analogs to scientific understandings of conventional notions of *internal validity* (labelled *credibility*), *external validity* (called *transferability*), *reliability* (named *dependability*), and *objectivity* (termed *neutrality*). *Credibility* pertaining to the truth value of research means that respondents' perceptions are faithfully represented in the research, while *transferability*, reflecting the applicability of qualitative research, calls for detailed description of the research context based on which the reader can judge possible similarities to their own. *Dependability* addresses the problem of consistency by calling for the trackability of any changes that occurred in the course of the research, and, finally, since true objectivity seems to be impossible even in physics (the dual nature of light is one example), the aspect of *neutrality* is ensured by confirmability of the data produced.

What seems to be the recent trend is that authors working within the qualitative paradigm attempt to break with those ideals of good research that are rooted in positivism and create their own criteria organically from the philosophical underpinnings of the qualitative research paradigm (cf. Creswell, 2018; Mirhosseini, 2020; Rallis & Rossman, 2009; Tracy, 2019). However, the existence of the many different qualitative research traditions such as phenomenology, narrative enquiry, ethnography or conversation analysis, to mention just a few, makes the creation of a unified framework similar to what exists in the positivist research tradition nearly impossible. As Lazaraton (2003) points it out "the nature of the research cannot be separated from the methods used to carry it out, which are implicated in the criteria used to judge it" (p. 9). One solution under these circumstances could be creating specific quality criteria unique to each research tradition. Another could be opting for the other extreme, creating very broad criteria that could synthesize different practices across theoretical traditions and paradigms. This latter approach is represented by Tracy's (2010, 2019) *eight big tent framework* for high quality qualitative research.

The *eight big tent framework* (Tracy, 2010, 2019) includes broad criteria, such as *worthy topic*, referring to the significance to the subject investigated, and *significant contribution*, addressing similar issues in connection with the research itself. *Rich rigor* emphasizes the care and effort taken to carry out the study

in an appropriate manner based on the theoretical constructs throughout the whole data collection and analysis process, while *sincerity* highlights the transparency of these issues and self-reflexivity about any values and biases. Similarly to the meaning proposed by Guba (1981), *credibility* aims to ensure that the reality represented in the research is plausible or appears to be true, whereas *resonance* refers to the effect that a study or its written account has on the audience. One of the final two tents house *ethical issues*, which, despite the fact that they are missing from Guba's (1981) original framework, feature increasingly prominently in more recent quality guidelines (Howitt, 2016; Mirhosseini, 2020; Rallis & Rossman, 2009). The other calls for *meaningful coherence*, which is a rather broad concept, and it reflects on the whole research process stating that qualitative studies should: "(a) achieve their stated purpose; (b) accomplish what they espouse to be about; (c) use methods and representation practices that partner well with espoused theories and paradigms; and (d) attentively interconnect literature reviewed with research foci, methods, and findings" (Tracy, 2010, p. 848).

Although the labels used for describing good quality research in qualitative studies clearly proliferate, there seems to be more common ground and agreement with regard to the steps or procedures that can be used to ensure that these ideals are met. Triangulation, thick description, member checks, peer debriefing, inter-coder reliability, audit trail, and ethical issues feature very frequently in the majority of works discussing quality control issues in qualitative research (see e.g., Creswell, 2018; Dörnyei, 2007; Guba, 1981; Mirhosseini, 2020; Patton, 2015; Rallies & Rossman, 2009; Tracy, 2019). Therefore, we would like to briefly explain what each of these concepts covers.

Triangulation is a concept that originates from navigational and land surveying techniques; it can be used to "determine a single point in space with the convergence of measurements taken from two other distinct points" (Rothbauer, 2008, p. 892). It is frequently used in social sciences, and usually four different types of triangulation are differentiated: triangulation of the methods of data collection, triangulation of data sources, investigator triangulation, and theory triangulation (Denzin, 1989). Although the original purpose of triangulation was the verification of the research findings by relying on data originating from different sources, it is now increasingly understood as a strategy that allows researchers to strengthen their findings and enrich their interpretations via the exploration of different perspectives, as the philosophical stance associated with qualitative research emphasizes uniqueness, which seems incompatible with the idea of establishing one objective truth.

Thick description is a procedure that can be used to provide rich detail about the different aspects of the research which is needed for more profound understanding of the context of the study and draws attention to the context-

embedded nature of qualitative research. *Member checks*, or *member reflections*, as Tracy (2019) chooses to refer to this procedure, are needed to ensure that the participants' viewpoints are represented in the study, which can be achieved by sharing the research data with the participants and asking their feedback on it. Both *peer debriefing* and *inter-coder reliability* mean involving fellow researchers in the research process either for the purpose of receiving overall feedback or for specifically ensuring the reliability of the coding process. Establishing an *audit trail* requires that different steps of the research process can be traced back later on, as these were documented throughout the study. Finally, careful consideration of *ethical issues* is expected to be present throughout the research process, and references to all of these need to appear to some extent in the write-up.

At an even more specific level, carrying out empirical research is not possible without data; therefore, regardless of the fact whether the study is quantitative or qualitative, data need to be collected. In some cases, researchers work with naturalistic data, some sort of artefacts that exist independently of researchers' endeavors; thus, they only need to be located and not created for the purpose of the research. Such data in our field is usually understood as some kind of text: books, diaries, blogs, lesson plans and so on. In the majority of cases, however, data are collected specifically for the purpose of the study; in these cases data collection tools are employed, which can be interviews, tasks, and tests, to name just a few. Observations occupy a middle ground here since the event observed might be external to the research (e.g., English lessons in a high school); nevertheless, the presence of a researcher or even just a recording device might alter the phenomenon under scrutiny. Although there are some data collection tools that are more clearly linked with the qualitative research tradition and are referred to as qualitative data collection tools, such as observations or interviews (see e.g., Dörnyei, 2007), the *method of data analysis* chosen is also determinant: for example, interviews can be analyzed quantitatively as well using content analysis (Fraenkel, et al., 2012).

Quality control considerations, which are present from the moment researchers start to plan their investigation to the point of write-up, are also relevant with regard to the *use* and the *reported use* – these are clearly different issues – of data collection tools. Although papers devoted to broader conceptual matters related to quality control considerations rarely address issues of this specificity, discussions aimed at the use of specific data collection tools might give us hints as to their proper application. Likewise, papers sharing guidelines on write-up might offer certain recommendations as to what needs to be reported or supplied in the form of supplementary materials (see e.g., Howitt, 2016).

With regard to *interviews* (Dörnyei, 2007; Fraenkel, et al., 2012; Howitt, 2016; Patton, 2015; Richards, 2009), it seems useful to indicate what type of interview

was used. Moreover, by sharing some details about the interview guide's content, the reader can be presented with a more informed picture about the topics discussed: including the interview guide in the appendix or some supplementary material, if space permits, would be a preferable solution here. Although interviews are expected to be recorded, information about this fact should also be given besides indicating the language of the interview, as the common assumption of conducting the interviews in the respondents' L1 might not always be fulfilled (Welch & Piekkari, 2006). Indicating the length of recordings or shedding light on the size of the resulting text corpus may give insights to the reader about the breadth and depth with which the topic might have been covered. Finally, in cases where several interviews are conducted, piloting the instrument is also an expected quality control step, as this could provide information about potential problems with the instrument before its application.

As regards *observations* (Dörnyei, 2007; Fraenkel, et al., 2012; Patton, 2015), recording the phenomenon observed may or may not make sense depending on circumstances. Moreover, the recording of events can render traditionally used data collection tools of observation like field notes, researcher diaries or observation grids unnecessary, as the event can be replayed countless times making fine-grained analyses possible. Of course, in the case of complex events involving many participants, like a language class, recording everything that happens in its entirety is close to impossible, so a combination of recordings and field notes might be a sensible choice. Data collection tools used in observations range from unstructured field notes to structured observation grids. In the case of the latter, a description of the main points to be observed or providing access to the instrument itself besides details about piloting are expected as quality control steps. Indicating the length of observation or the size of the resulting text corpus in some form is also helpful for the reader in judging the volume of the data.

Despite their growing popularity as data collection tools, *tasks* are less established instruments than interviews or observations, which is evident from that fact that while research methodology handbooks all contain chapters on interviews and observation, tasks are hardly ever mentioned although they probably fit the broad category of elicited learner language (Mirhosseini, 2020). Nevertheless, recording and piloting as well as providing a brief description of content or even sample tasks in the text or supplementary materials seem to be justifiable expectations in connection with them as well.

3. Research questions

Since the number of qualitative research articles seems to be on the rise when it comes to investigating IDs in language learning, providing a systematic review

of such articles that have appeared in leading journals of our field is timely. We have intentionally selected top-tier journals as they present the leading voices and often cited articles, thus creating and maintaining quality-control-related expectations. A systematic review allows offering insights into the distribution of qualitative and quantitative articles across the journals and the main topics investigated in them. It can also shed light on the most popular data collection instruments and those quality control steps that were applied in connection with them to ensure that they meet the high standards set by these journals. In order to achieve these aims, we formulated the following research questions:

1. What are the distributional characteristics of research studies on individual differences reported in top-tier journals of applied linguistics in recent years? (RQ1)
2. What are the main topics of qualitative studies on individual differences reported in top-tier journals of applied linguistics in recent years? (RQ2)
3. What data collection tools are employed in qualitative studies on individual differences reported in top-tier journals of applied linguistics in recent years? (RQ3)
4. What is reported in terms of quality control in connection with three frequently used data collection tools (interviews, observations, and tasks) in qualitative studies on individual differences published in top-tier journals of applied linguistics in recent years? (RQ4)

4. Research methods

4.1. Criteria for inclusion: The journals

As we had planned to investigate quality-control issues, we decided to focus on top-tier journals in our study to identify selective and high-quality publishing policies. The selection of journals to be included in our analysis was based on the following process. First, the scientific journal ranking (SJR) of the journals in the category of “language and linguistics” was downloaded from the Scimago Institution’s homepage <https://www.scimagojr.com/journalrank.php?category=1203>. Subsequently, for the first 20 journals on the list, two further indices were looked up. The 2019 impact factor of the journals was copied from the 2020 Edition of the Journal Citation Reports® (JCR) published by Clarivate Analytics <https://www.annualreviews.org/page/librarians/impact-factors> and they were rank ordered based on this index as well. Finally, the so-called SNIP (source normalized impact per publication) score of the journals was also established based on the CWTS website <https://www.journalindicators.com/indicators>, and a third

rank order was prepared for the journals based on this. The three rank orders were then added up, creating a final order for the journals (see Table 1). Before creating the final list of journals, those not dealing with topics related to second or foreign language learning or teaching were eliminated from the list, along with those journals where all three indices described above were not available. We then arrived at the following list of journals: 1. *Applied Linguistics*, 2. *Modern Language Journal*, 3. *Studies in Second Language Acquisition*, 4. *Language Teaching Research*, 5. *Language Learning*, 6. *Journal of Second Language Writing*, 7. *Language Teaching*, 8. *Language Learning and Technology*, 9. *Computer Assisted Language Learning*, and 10. *TESOL Quarterly*. In our search, we decided to start looking for relevant articles in the first five journals, including the past 5 years, ranging from 2016-2020.

Table 1 The list of journals with the final rank order indicated by Roman numbers

Title	SJR (rank)	SNIP source normalized impact per publication (rank)	2019 impact factor (rank)	ranks added	final rank order
<i>Modern Language Journal</i>	3.966 (1)	3.43 (2)	3.538 (7)	10	II.
<i>Journal of Communication</i>	3.259 (2)	2.88 (7)	4.842 (1)	10	
<i>Studies in Second Language Acquisition</i>	3.238 (3)	2.76 (9)	2.838 (11)	23	III.
<i>Applied Linguistics</i>	3.148 (4)	4.36 (1)	4.286 (2)	7	I.
<i>Journal of Memory and Language</i>	3.093 (5)	2.19 (13)	3.893 (3)	21	
<i>Language Teaching Research</i>	2.596 (6)	3.11 (5)	2.647 (13)	24	IV.
<i>Communication Research</i>	2.410 (7)	2.69 (10)	3.758 (5)	22	
<i>Language Learning</i>	2.404 (8)	2.38 (11)	3.408 (8)	27	V.
<i>Cognition</i>	2.330 (9)	1.82 (17)	3.294 (9)	35	
<i>Communication Theory</i>	2.185 (10)	2.91 (6)	1.905 (18)	34	
<i>Journal of Experimental Psychology</i>	2.158 (11)	1.52 (19)	2.832 (12)	42	
<i>Language Learning and Technology</i>	2.011 (12)	2.09 (15)	2.473 (15)	42	VIII.
<i>Annual Review of Applied Linguistics</i>	1.979 (13)	-	1.679 (19)	32+	
<i>Communication Monographs</i>	1.979 (14)	3.12 (4)	3.843 (4)	22	
<i>Journal of Second Language Writing</i>	1.963 (15)	3.31 (3)	3.077 (10)	28	VI.
<i>Language Teaching</i>	1.867 (16)	2.80 (8)	3.714 (6)	30	VII.
<i>Computer Assisted Language Learning</i>	1.848 (17)	2.31 (12)	2.642 (14)	43	IX.
<i>CALICO Journal</i>	1.827 (18)	1.72 (18)	-	36+	
<i>Journal of Literacy Research</i>	1.813 (19)	1.91 (16)	2.255 (16)	51	
<i>TESOL Quarterly</i>	1.725 (20)	2.16 (14)	2.071 (17)	51	X.

4.2. Criteria for inclusion: The articles

In this phase of our study, we independently selected articles from the journals by first including those that investigated issues in individual differences research between the years 2016 and 2020. We decided to cover a five-year span in order to have a fairly large number but relatively recent studies to analyze. We approached individual difference variables in the broadest sense and defined them for the purpose of the selection as any variable aiming to measure differences

among learners that might impact learning processes or outcomes. Once we agreed on the final number of articles ($N = 371$), the articles were categorized into five groups: quantitative, mono- or multi-methods qualitative, mixed methods with qualitative parts and finally studies using other methods (e.g., meta- or theoretical analysis). We included articles from special issues but excluded short communications. In the analysis, we worked with studies containing mono- or multi-methods qualitative methods and mixed methods with qualitative parts. Thus, our final sample consisted of 93 research articles.

4.3. Coding and analysis

When coding the information in the selected articles, we employed a cyclical coding process. First, basic information about the selected articles was recorded, such as the author(s)' name, the title of the article and the year of publication. With regard to the publication date, we used the volumes from the journals' websites and checked volumes published between 2016 and 2020, disregarding information about online-first publications. In this round, we also recorded the main ID topic of the articles along with the data collection instruments applied in them (RQ1, RQ2). This enabled us to further divide our 93 qualitative or partially qualitative articles into the following three categories: (1) mono-method qualitative studies, (2) multi-method qualitative studies, and (3) mixed methods studies containing a qualitative part RQ 3. In the next round of coding, the data collection tools used in the selected qualitative articles were recorded. Finally, we examined quality control steps in connection with two popular qualitative instruments, interviews and observations, and a fairly frequently used third instrument, language tasks (RQ4). The coding in this case was based on the following five common categories: (1) mentioning audio or video recording, (2) reference to length or corpus size, (3) description about the content of the instrument used, (4) whether the instrument (or a sample of it) was provided either in the text or in the appendix, and (5) mention of any attempt at piloting, that is, trying out the instrument. In the case of observations, we also recorded whether a tool (e.g., field notes or an observation grid) was used for data collection as well; in this case, the description of content and piloting obviously referred to this instrument. In the case of interviews, we noted if the language of the interview was mentioned, while in the case of tasks we recorded whether the task was a receptive or productive one. For productive tasks we indicated the modality (oral or written) and took note of the task type as well. In order to enhance the quality of our study, both of us coded all the articles, results were compared and differences discussed and agreed upon (for the coding schemes applied in various phases of the research see Appendix).

5. Results and discussion

In this part of the article, we will present and discuss our research results according to our research questions. We start with the distributional characteristics of the articles in the sample. Next, the topical and methodological analyses are presented. Finally, we deal with the issues pertaining to quality control in the articles.

5.1. Distributional characteristics of research studies on individual differences reported in top-tier journals of applied linguistics in recent years

Table 2 displays the distributional results across the five journals and information about the chosen research approach in the articles. The results show great variation among the journals and there is a clear rank order concerning the number of articles focusing on individual differences research with *Studies in Second Language Acquisition* having the top position, while *Applied Linguistics* containing the fewest articles on the topic. If we take a look at the research approach of the journals, it is clear that the majority of the ID research studies employed quantitative research methods and only a small proportion of empirical investigations employed exclusively qualitative methods or mixed approaches with qualitative parts. There are only two journals, *Modern Language Journal* and *Language Teaching Research*, in which the proportions of quantitative and qualitative studies were comparable. These results are somewhat in contrast with the encouraging tendency to rely more and more often on qualitative studies in one particular ID field, that is, L2 motivation (Dörnyei & Ryan, 2015). Moreover, mixed methods studies clearly dominate our sample, as 47 out of the 93 articles, that is, over 50%, were mixed methods studies containing a quantitative component as well. The remaining 46 articles were equally divided between mono-method qualitative (23 articles) and multi-methods qualitative (23 articles) studies, suggesting that around 25% of all the studies included in our database relied on a single data source.

Table 2 Information on the articles in the database (2016-2020)

Journal (N)	ID theme	Research approach				
		Quantitative ¹	Mono method qualitative	Multi method qualitative	Mixed methods study	Other ²
<i>Studies in Second Language Acquisition</i> (N = 171)	100	85	1	0	6	8
<i>Language Learning</i> (N = 195)	93	81	0	0	1	11
<i>Modern Language Journal</i> (N = 183)	84	35	15	10	15	9
<i>Language Teaching Research</i> (N = 148)	55	27	2	6	20	0
<i>Applied Linguistics</i> (N = 174)	46	26	5	7	5	3

Note. ¹Quantitative study using either one or multiple data collection instruments. ²Theoretical or meta-analytic studies

5.2. Main topics in qualitative studies and mixed methods studies

It is clear from Table 3 that there are a number of topics that are probably more likely to be targeted from a qualitative perspective or by using at least partially qualitative methods. These topics are motivation, various cognitive processes influencing language acquisition, identity, language learning experiences, beliefs about language learning, and strategies. Since motivation and strategies are constructs that have been extensively researched with the help of quantitative methods in the past, in their case the use of qualitative methods represents a novel approach with hopefully new insights. By contrast, identity, language learning experiences and beliefs are fairly complex issues whose investigation has involved qualitative methods for quite some time. The relatively large number of studies devoted to analyzing different cognitive processes, such as attention, noticing, explicit and implicit learning, and awareness, involved in language acquisition might represent the broadening of the research agenda through a closer examination of cognitive processes, where qualitative measures are typically used to explain findings derived from traditional, quantitative ones.

As Table 3 only contains the main topics identified, it is also important to point out that only a minority of the articles ($N = 16$) focused on more than one topic. Some examples of studies addressing multiple IDs include identities and experiences (Anderson, 2019; Brown, 2016), motivation, emotions and beliefs (e.g., Csizér & Kontra, 2020; Poupore, 2018) as well as perceptions, beliefs and emotions (e.g., Jung & Révész, 2018; Kormos & Préfontaine, 2017). It seems that the common perception that qualitative studies target multiple IDs within the same research design cannot be supported with these results.

Table 3 Main topics identified in the articles between 2016 and 2020

Main topics	Number of articles (%)
Motivation	24 (26)
Cognitive processes involved in language acquisition (attention, noticing, explicit and implicit learning, awareness, etc.)	21 (23)
Identity	12 (13)
Language learning experiences (e.g. study/work abroad, in the classroom)	11 (12)
Beliefs about language learning	9 (10)
Strategies	6 (6)
Willingness to communicate/communication style	3 (3)
Self-efficacy	2 (2)
Age	2 (2)
Group dynamics	1 (1)
Personality	1 (1)
Emotions	1 (1)
Total	93 (100%)

5.3. Data collection tools used in qualitative and mixed methods studies

The 93 articles in our database used 223 data collection instruments in total (see Table 4), which reflects a trend that the majority of the studies reported in the articles used more than one data collection tool. When we coded the different data collection tools used in qualitative and partially qualitative mixed methods studies, we took all data collection tools appearing in the articles into consideration. The reason for this decision is that sometimes it is quite hard to establish whether a certain data collection instrument was only employed to collect data for the qualitative part of the study or not, as is the case when biographical data are collected with the help of questionnaires and then are also used when interpreting data collected for the qualitative phase. This is the background against which our results should be interpreted.

Table 4 Data collection tools found in the articles between 2016 and 2020

Type of data collected	No. of occurrences	Data collection tool	No. of occurrences
Spoken or written language produced in response to some prompt or question	99 (44%)	Interview	55 (25%)
		Focus group interview	9 (4%)
		Stimulated recall	13 (6%)
		Think aloud protocol	4 (2%)
		Reflection	5 (2%)
		Diary or journal	4 (2%)
		Other type of spoken or written language produced in response to some prompt or question (e.g., narrative, verbal report, identity memo, etc.)	9 (4%)
Non-elicited, naturalistic spoken data	35 (16%)	Observation/recording	35 (16%)
Spoken or written quantifiable data	76 (34%)	Language tasks	20 (9%)
		Language tests	12 (5%)
		Questionnaires	33 (15%)
		Other measurements, ratings, scores or logs	11 (5%)
Non-elicited, naturalistic written data (no data collection tool)	13 (6%)	Naturally occurring texts (e.g., emails, books, blogs, etc.)	5 (2%)
		Teaching materials (e.g., lesson plans, syllabus, course book, etc.)	8 (3%)
Total	223 (100%)		223 (100%)

In light of the above, the relatively high percentage of questionnaires (15%), language tests (5%) and measurement data (5%) is perhaps not surprising. As expected, the largest proportion of data collection tools comprised interviews (25%) which, together with other data collection tools aimed at eliciting spoken or written language in response to some question or prompt, made up almost half of all the data collection tools (44%) found in the articles. A relatively smaller proportion of data were collected with the help of observations, accounting for 16% of the instruments, while naturalistic written data were only collected in 6% of the studies in question. Language tasks which often lead to

numerically quantifiable measures made up 9% of all the data collection instruments identified, which raises the question as to why this type of instrument is not discussed more prominently in publications concerned with conducting research in applied linguistics.

5.4. Quality control measures reported in connection with interviews, observations and language tasks

In terms of the six quality control measures we investigated in relation to the qualitative interview guides used in the studies (1 – recording, 2 – length, 3 – content, 4 – sample instrument, 5 – piloting, and 6 – language), we can report a mixed picture. As for the actual content of the interviews, we used three codes depending on the amount of information included in the article, appendix or as supplementary information. Out of the 55 articles, 14 contained hardly any information on the content of the interview, while 10 included the instrument as an attachment (nine as appendix, one in the supplementary information package) or detailed information was given in the article. The majority of the remaining articles contained some information about questions or prompts used in the study; thus, the reader could form an impressionistic view of the data collected. Concerning the piloting of the instrument, only four studies included some insight into this process either by stating that the instrument was tried out before use or by commenting on some elements of the pilot process. This piece of information is difficult to process because piloting instruments before use seems to be a generally advocated guideline (Dörnyei, 2007; Howitt, 2016; Richards, 2009).

Another important issue is the language of the data collection. According to Welch and Piekkari (2006), this issue is rarely discussed explicitly in interview studies probably due to the fact that, based on common assumptions, interviews are conducted in the participants' L1. In the articles analyzed, we found a roughly equal number of studies that reported information on the language used ($N = 27$) and those that did not ($N = 28$). When the information was not given, we very often had the impression that the language of the data collection was in fact the L1 of the participants, which was probably the intention of the researchers. However, for obvious reasons, we cannot be sure of this without the authors explicitly reporting this piece of information. In view of these considerations, when the focus of a piece of research is on some aspect of language learning, and participants are typically learners who can be considered bilingual to some extent, providing information about whether they were proficient enough in the language of the interview to be able to express their views clearly appears quite crucial. The last two categories coded concerned the length of the interview, which provides information about the volume of data collected, as well as whether the interviews were recorded. Based

on our analysis, 30 articles contained satisfactory information about the lengths of the interviews and thus gave the reader an understanding about the volume of data collected, while 25 articles reported no such information. As for data recording, we did not find the information in 11 articles. Again, this does not mean that the data was not recorded but simply that it was not reported explicitly.

Data recording was a very popular option in the observation studied reviewed: 26 out of the 34 studies used either video or audio recording or a combination of the two. Field or observation notes, however, appeared to be less popular as they were only used in 13 articles, while structured observation instruments were employed even less frequently. They were only mentioned in four articles along with research journals used in two studies. In seven out of the eight articles where no audio or video recording was used, the application of some form of note taking was mentioned, leaving one study where no data recording was reported whatsoever. Quite understandably, providing information about the content of the data recording tool was only meaningful in the case of structured instruments; in three out of the four cases such information or the instrument itself was made available for the readers. Piloting would have only made sense in connection with these four instruments; however, no reference to piloting could be found in any of the articles. The fact the three out of the four instruments were based on either COLT (Spada & Frolich, 1995) or MOLT (Guilloteaux & Dörnyei, 2008) might be responsible for this finding, as the authors might have considered the piloting process unnecessary. The length of observation either per occasion or in total was recorded in 26 cases while no such information was provided in eight studies.

As regards tasks, it was possible to establish for all 20 articles whether they used productive/output tasks ($N = 15$), receptive/input tasks ($N = 3$) or a combination of the two ($N = 2$). Out of the 17 articles containing output tasks, three required written production, 13 oral production and one both. Out of the 14 cases when oral production tasks were used, either the audio or video recording of the data was clearly indicated in 11 cases, it could be implied in two, and there was only one article where it was unclear whether oral performance was recorded or not. Task descriptions were in some detail provided in 18 out of the 20 articles; in those two where such information was not given, the authors reported using a variety of different tasks. In eight cases out of the 18, sample tasks or task instructions were provided either in the article itself or in some form of supplementary material. Reference to the piloting of the tasks, however, was only made in three cases. The length of individual tasks or of the whole corpus was provided in 11 cases. A label for the task type was provided in all 20 articles; nevertheless, the great variety of existing task types (cf. Ellis, 2018) makes such labels moderately useful. The most popular task types were narratives used in five cases, opinion gap tasks used in three and description, argumentation, and information gap tasks, each used in two cases.

6. Conclusion

Based on our systematic review of the 93 fully or partially qualitative research articles dealing with the topics of ID factors appearing in top tier journals of applied linguistics in the past 5 years, the following answers can be provided to our research questions. As regards the distribution of articles across these journals, certain tendencies can be identified. As far as the investigation of ID variables is concerned, *Studies in Second Language Acquisition* and *Language Learning* mostly tend to publish numerous, mainly quantitative articles on this topic, while in *Modern Language Journal* and *Language Teaching Research* we found a larger proportion of studies taking at least a partially qualitative approach in this respect. With regard to the most popular topics, a few, well-established ID variables such as motivation, beliefs and strategies seem to be the dominant theme in many articles, with research on emerging topics such as identity and learning experiences also gaining ground. We also identified a diverse group of articles dealing with information processing involved in language acquisition. These studies mainly adopted mixed methods designs, which might indicate a shift of focus in the field of research on IDs.

The most frequently used data collection tools reported in the articles turned out to be interviews, followed by observations, questionnaires and various language tasks. As the large number of data collection tools indicate, mono-method studies were relatively rare among the articles reviewed while the heavy reliance on questionnaires can probably be ascribed to the overrepresentation of mixed methods studies in the corpus. Quality control issues were investigated in connection with the data collection instruments, and we found that in certain cases important quality control steps were either ignored or simply not reported. The most shocking finding in this regard was the low number of articles reporting previous piloting of their data collection tools. As for the overall quality control issues presented in the literature review, we think that it is not only the quality of data collection instruments that should be considered but also overarching issues such as triangulation, thick description, member checks, peer debriefing, inter-coder reliability, audit trail, and ethical issues. Unfortunately, including all of these issues in this article would have been impossible.

Our systematic review has certain limitations primarily linked to the depth of analysis concerning quality control issues. Since we firmly believe that quality control should be present from the start of the design until the point of write-up, we feel that there are a number of other relevant issues that should have been checked in the papers; however, space limitations prevented us from carrying out a more in-depth review in this respect. Therefore, investigating a larger number of potential quality control issues in this corpus will be a task for future

researchers. Another limitation is that we selected top-tier journals in order to include the highest quality research in this study. We understand that our sampling presents an inflated view of the quality of the studies in the field, but this was done on purpose to see what the highest-level publications have to offer in terms of quality control. Future studies, though, need to consider other journals, monographs as well as unpublished sources, such as PhD dissertations.

Acknowledgments

This study was funded by the Research Program for Public Education Development of the Hungarian Academy of Sciences. The authors are members of the MTA-ELTE Foreign Language Teaching Research Group.

References¹

- *Acheson, K., Taylor, J., & Luna, K. (2016). The burnout spiral: The emotion labor of five rural US foreign language teachers. *Modern Language Journal, 100*(2), 522-537. <https://doi.org/10.1111/modl.12333>
- Albert, Á., Dóczi, B., Csizér, K., & Piniel, K. (in press). The contribution of motivation and emotions to language learning autonomy in the Hungarian secondary school classroom: The results of a questionnaire study. In V. De Wilde & C. Goriot (Eds.), *Individual differences in young second language learners*. De Gruyter.
- Albert, Á., & Piniel, K. (2021). Changes in students' motivation, autonomy, self-efficacy, and emotions across a school year: Examining possible effects of switching to online education. In G. Tankó & K. Csizér (Eds.), *DEAL 2021: Current explorations in English applied linguistics* (pp. 265-298). Eötvös Loránd University.
- *Anderson, T. (2019). Reproductions of Chinese transnationalism: Ambivalent identities in study abroad. *Applied Linguistics, 40*(2), 228-247. <https://doi.org/10.1093/applin/amx018>
- *Back, M. (2016). Epistemics and expertise in peer tutoring interactions: Co-constructing knowledge of Spanish. *Modern Language Journal, 100*(2), 508-521. <https://doi.org/10.1111/modl.12334>
- *Barkhuizen, G. (2017). Investigating language tutor social inclusion identities. *Modern Language Journal, 101*(S1), 61-75. <https://doi.org/10.1111/modl.12369>
- *Bell, P. K. (2017). Explicit and implicit learning: Exploring their simultaneity and immediate effectiveness. *Applied Linguistics, 38*(3), 297-317. <https://doi.org/10.1093/applin/amv028>
- *Brown, L. (2016). An activity-theoretic study of agency and identity in the study abroad experiences of a lesbian nontraditional learner of Korean. *Applied Linguistics, 37*(6), 808-827. <https://doi.org/10.1093/applin/amu075>
- *Burri, M., Chen, H., & Baker, A. (2017). Joint development of teacher cognition and identity through learning to teach L2 pronunciation. *Modern Language Journal, 101*(1), 128-142. <https://doi.org/10.1111/modl.12388>
- *Busse, V. (2017). Plurilingualism in Europe: Exploring attitudes toward English and other European languages among adolescents in Bulgaria, Germany, the Netherlands, and Spain. *Modern Language Journal, 101*(3), 566-582. <https://doi.org/10.1111/modl.12415>
- *Butler, Y. G. (2017). Motivational elements of digital instructional games: A study of young L2 learners' game designs. *Language Teaching Research, 21*(6), 735-750. <https://doi.org/10.1177/1362168816683560>

¹ References marked with an asterisk indicate studies included in the systematic review.

- Carroll, J. M., & Sapon, S. M. (1959). *Modern language aptitude test*. Psychological Corporation.
- * Caras, A. S. L., & Leow, R. P. (2016). The effectiveness of guided induction versus deductive instruction on the development of complex Spanish gustar structures: An analysis of learning outcomes and processes. *Studies in Second Language Acquisition*, 38(2), 265-291. <https://doi.org/10.1017/S0272263116000139>
- * Charalambous, P., Charalambous, C., & Rampton, B. (2017). De-securitizing Turkish: Teaching the language of a former enemy, and intercultural language education. *Applied Linguistics*, 38(6), 800-823. <https://doi.org/10.1093/applin/amv063>
- * Chen, Q., & Wright, C. (2017). Contextualization and authenticity in TBLT: Voices from Chinese classrooms. *Language Teaching Research*, 21(4), 517-538. <https://doi.org/10.1177/1362168816639985>
- * Chen, Y. S. (2016). Understanding the development of Chinese EFL learners' email literacy through Exploratory Practice. *Language Teaching Research*, 20(2), 165-180. <https://doi.org/10.1177/1362168814562014>
- Chong, S. W., & Plonsky, L. (2021a). A primer on qualitative research synthesis in TESOL. *TESOL Quarterly*, 55, 1024-1034. <https://doi.org/10.1002/tesq.3030>
- Chong, S. W., & Plonsky, L. (2021b, June 15). A typology of secondary research in Applied Linguistics. <https://doi.org/10.31219/osf.io/msjrh>
- * Coffey, S. (2018). Choosing to study modern foreign languages: Discourses of value as forms of cultural capital. *Applied Linguistics*, 39(4), 462-480. <https://doi.org/10.1093/applin/amw019>
- Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Prentice Hall.
- Creswell, J. W. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed). Sage.
- Csikszentmihályi, M., Abuhamdeh, S., & Nakamura, J. (2005). Flow. In A. Elliot & C. Dweck (Eds.), *Handbook of competence and motivation* (pp. 598-623). The Guilford Press.
- Csizér, K., & Albert, Á. (2021). Trait and state perspectives of individual difference research. In S. Mercer & T. Gregersen (Eds.), *The Routledge handbook of psychology of language learning* (pp. 339-349). Routledge.
- Csizér, K., Albert, Á., & Piniel, K. (2021). The interrelationship of language learning autonomy, self-efficacy, motivation and emotions: The investigation of Hungarian secondary school students. In M. Pawlak (Ed.), *Investigating individual learner differences in second language learning* (pp. 1-23). Springer.
- *Csizér, K., & Kontra, E. H. (2020). Foreign language learning characteristics of deaf and severely hard-of-hearing students. *Modern Language Journal*, 104(1), 233-249. <https://doi.org/10.1111/modl.12630>

- *Culpeper, J., & Kan, Q. (2020). Communicative styles, rapport, and student engagement: An online peer mentoring scheme. *Applied Linguistics*, 41(5), 756-786. <https://doi.org/10.1093/applin/amz035>
- *Dávila, L. T. (2019). "J'aime to be funny!:" Humor, learning, and identity construction in high school English as a second language classrooms. *Modern Language Journal*, 103(2), 502-514. <https://doi.org/10.1111/modl.12557>
- Delamont, S. (Ed.). (2012). *Handbook of qualitative research in education*. Edward Elgar Publishing.
- Denzin, N. K. (1989). *The research act: A theoretical introduction to sociological methods*. Prentice Hall.
- Dewaele, J.-M., & MacIntyre, P. D. (2016). Foreign language enjoyment and foreign language classroom anxiety: The right and left feet of FL learning? In P. D. MacIntyre, T. Gregersen, & S. Mercer (Eds.), *Positive psychology in SLA* (pp. 215-236). Multilingual Matters.
- *Donaghue, H. (2020). Teachers and supervisors negotiating identities of experience and power in feedback talk. *Modern Language Journal*, 104(2), 401-417. <https://doi.org/10.1111/modl.12633>
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.
- Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. Routledge.
- Dörnyei, Z., & Ushioda, E. (2011). *Teaching and researching motivation* (2nd ed.). Longman.
- Ellis, R. (2018). *Reflections on task-based teaching*. Multilingual Matters.
- *Ender, A. (2016). Implicit and explicit cognitive processes in incidental vocabulary acquisition. *Applied Linguistics*, 37(4), 536-560. <https://doi.org/10.1093/applin/amu051>
- *Fang, F., & Baker, W. (2018). 'A more inclusive mind towards the world': English language teaching and study abroad in China from intercultural citizenship and English as a lingua franca perspectives. *Language Teaching Research*, 22(5), 608-624. <https://doi.org/10.1177/1362168817718574>
- Flick, U. (2018). *Designing qualitative research*. Sage.
- Fraenkel, J. R., Wallen, N.E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill Inc.
- *Fukuta, J. (2016). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research*, 20(3), 321-340. <https://doi.org/10.1177/1362168815570142>
- *Gkonou, C., & Miller, E. R. (2019). Caring and emotional labor: Language teachers' engagement with anxious learners in private language school classrooms.

Language Teaching Research, 23(3), 372-387. <https://doi.org/10.1177/1362168817728739>

- *Godfroid, A. (2016). The effects of implicit instruction on implicit and explicit knowledge development. *Studies in Second Language Acquisition*, 38(2), 177-215. <https://doi.org/10.1017/S0272263115000388>
- *Graus, J., & Coppen, P. A. (2017). The interface between student teacher grammar cognitions and learner-oriented cognitions. *Modern Language Journal*, 101(4), 643-668. <https://doi.org/10.1111/modl.12427>
- *Graus, J., & Coppen, P. A. (2018). Influencing student teacher grammar cognitions: The case of the incongruous curriculum. *Modern Language Journal*, 102(4), 693-712. <https://doi.org/10.1111/modl.12510>
- *Gu, M. M., & Canagarajah, S. (2018). Harnessing the professional value of a transnational disposition: Perceptions of migrant English language teachers in Hong Kong. *Applied Linguistics*, 39(5), 718-740. <https://doi.org/10.1093/applin/amw048>
- Guba, E. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal*, 29(2), 75-91. <https://doi.org/10.1007/BF02766777>
- Guilloteaux, M. J., & Dörnyei, Z. (2008). Motivating language learners: A classroom-oriented investigation of the effects of motivational strategies on student motivation. *TESOL Quarterly*, 42(1), 55-77. <https://doi.org/10.1002/j.1545-7249.2008.tb00207.x>
- *Gurzynski-Weiss, L. (2016). Factors influencing Spanish instructors' in-class feedback decisions. *Modern Language Journal*, 100(1), 255-275. <https://doi.org/10.1111/modl.12314>
- Heigham, J. & Croker, R. (Eds.). (2009). *Qualitative research in applied linguistics: A practical introduction*. Palgrave MacMillan.
- *Henry, A., & Davydenko, S. (2020). Thriving? Or surviving? An approach-avoidance perspective on adult language learners' motivation. *Modern Language Journal*, 104(2), 363-380. <https://doi.org/10.1111/modl.12635>
- *Henry, A., & Thorsen, C. (2018). Teacher-student relationships and L2 motivation. *Modern Language Journal*, 102(1), 218-241. <https://doi.org/10.1111/modl.12446>
- *Henry, A., & Thorsen, C. (2020). Disaffection and agentic engagement: "Redesigning" activities to enable authentic self-expression. *Language Teaching Research*, 24(4), 456-475. <https://doi.org/10.1177/1362168818795976>
- *Higgins, C., & Ponte, E. (2017). Legitimizing multilingual teacher identities in the mainstream classroom. *Modern Language Journal*, 101(S1), 15-28. <https://doi.org/10.1111/modl.12372>

- *Hiver, P. (2017). Tracing the signature dynamics of language teacher immunity: A retrodictive qualitative modeling study. *Modern Language Journal*, 101(4), 669-690. <https://doi.org/10.1111/modl.12433>
- *Hlas, A. C., Neyers, K., & Molitor, S. (2019). Measuring student attention in the second language classroom. *Language Teaching Research*, 23(1), 107-125. <https://doi.org/10.1177/1362168817713766>
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal*, 70, 125-132.
- Howitt, D. (2016). *Introduction to qualitative research methods in psychology* (3rd ed.). Pearson.
- *Jackson, D. O., & Cho, M. (2018). Language teacher noticing: A socio-cognitive window on classroom realities. *Language Teaching Research*, 22(1), 29-46. <https://doi.org/10.1177/1362168816663754>
- *Jing-Schmidt, Z., Zhang, Z., & Chen, J.-Y. (2016). Identity development in the ancestral homeland: A Chinese heritage perspective. *Modern Language Journal*, 100(4), 797-812. <https://doi.org/10.1111/modl.12348>
- *Jung, J., & Révész, A. J. (2018). The effects of reading activity characteristics on L2 reading processes and noticing of glossed constructions. *Studies in Second Language Acquisition*, 40(4), 755-780. <https://doi.org/10.1017/S0272263118000165>
- *Jung, Y., Kim, Y., Lee, H., Cathey, R., Carver, J., & Skalicky, S. (2019). Learner perception of multimodal synchronous computer-mediated communication in foreign language classrooms. *Language Teaching Research*, 23(3), 287-309. <https://doi.org/10.1177/1362168817731910>
- *Kamiya, N. (2018). The effect of learner age on the interpretation of the non-verbal behaviors of teachers and other students in identifying questions in the L2 classroom. *Language Teaching Research*, 22(1), 47-64. <https://doi.org/10.1177/1362168816658303>
- *Kartchava, E., Gatlinton, E., Ammar, A., & Trofimovich, P. (2020). Oral corrective feedback: Pre-service English as a second language teachers' beliefs and practices. *Language Teaching Research*, 24(2), 220-249. <https://doi.org/10.1177/1362168818787546>
- *Kibler, A. K., Karam, F. J., Futch Ehrlich, V. A., Bergey, R., Wang, C., & Molloy Elreda, L. (2018). Who are 'long-term English learners'? Using classroom interactions to deconstruct a manufactured learner label. *Applied Linguistics*, 39(5), 741-765. <https://doi.org/10.1093/applin/amw039>
- *Kim, J., & Nassaji, H. (2018). Incidental focus on form and the role of learner extraversion. *Language Teaching Research*, 22(6), 698-718. <https://doi.org/10.1177/1362168817699642>
- Kormos, J., & Cszízér, K. (2014). Longitudinal changes in the interaction of motivation and inter-cultural contact in a study abroad context. In J. Horváth

& P. Medgyes (Eds), *Studies in honor of Marianne Nikolov* (pp. 9-22). Lingua Franca Csoport.

- *Kormos, J., & Préfontaine, Y. (2017). Affective factors influencing fluent performance: French learners' appraisals of second language speech tasks. *Language Teaching Research*, 21(6), 699-716. <https://doi.org/10.1177/1362168816683562>
- *Kremmel, B., Brunfaut, T., & Alderson, J. C. (2017). Exploring the role of phraseological knowledge in foreign language reading. *Applied Linguistics*, 38(6), 848-870. <https://doi.org/10.1093/applin/amv070>
- *Krulatz, A., Steen-Olsen, T., & Torgersen, E. (2018). Towards critical cultural and linguistic awareness in language classrooms in Norway: Fostering respect for diversity through identity texts. *Language Teaching Research*, 22(5), 552-569. <https://doi.org/10.1177/1362168817718572>
- *Lambert, C., & Zhang, G. (2019). Engagement in the use of English and Chinese as foreign languages: The role of learner-generated content in instructional task design. *Modern Language Journal*, 103(2), 391-411. <https://doi.org/10.1111/modl.12560>
- Lazaraton, A. (2003). Evaluative criteria for qualitative research in applied linguistics: Whose criteria and whose research? *Modern Language Journal*, 87(1), 1-12.
- *Lee, I., Mak, P., & Burns, A. (2016). EFL teachers' attempts at feedback innovation in the writing classroom. *Language Teaching Research*, 20(2), 248-269. <https://doi.org/10.1177/1362168815581007>
- *Lee, M. K., & Evans, M. (2019). Investigating the operating mechanisms of the sources of L2 writing self-efficacy at the stages of giving and receiving peer feedback. *Modern Language Journal*, 103(4), 831-847. <https://doi.org/10.1111/modl.12598>
- *Lee, S. H., & Kinginger, C. (2018). Narrative remembering of intercultural encounters: A case study of language program reintegration after study abroad. *Modern Language Journal*, 102(3), 578-593. <https://doi.org/10.1111/modl.12505>
- Long, M. (2014). *Second language acquisition and task-based language teaching*. John Wiley & Sons.
- *López-Serrano, S., de Larios, J. R., & Manchón, R. M. (2019). Language reflection fostered by individual L2 writing tasks. *Studies in Second Language Acquisition*, 41(3), 503-527. <https://doi.org/10.1017/S0272263119000275>
- *Louw, S., Watson Todd, R., & Jimarkon, P. (2016). Teacher trainers' beliefs about feedback on teaching practice: Negotiating the tensions between authoritativeness and dialogic space. *Applied Linguistics*, 37(6), 745-764. <https://doi.org/10.1093/applin/amu062>

- Lowie, W. M., & Verspoor, M. H. (2019). Individual differences and the ergodicity problem. *Language Learning*, 69(S1), 184-206.
- *Macaro, E., Tian, L., & Chu, L. (2020). First and second language use in English medium instruction contexts. *Language Teaching Research*, 24(3), 382-402. <https://doi.org/10.1177/1362168818783231>
- *MacIntyre, P. D., Baker, S. C., & Sparling, H. (2017). Heritage passions, heritage convictions, and the rooted L2 self: Music and Gaelic language learning in Cape Breton, Nova Scotia. *Modern Language Journal*, 101(3), 501-516. <https://doi.org/10.1111/modl.12417>
- MacIntyre, P. D., Burns, C., & Jessome, A. (2011). Ambivalence about communicating in a second language: A qualitative study of French immersion students' willingness to communicate. *Modern Language Journal*, 95, 81-96.
- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *Modern Language Journal*, 82, 545-562.
- MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, 44, 283-305.
- MacIntyre, P. D., & Legatto, J. (2011). A dynamic system approach to willingness to communicate: Developing an idiodynamic method to capture rapidly changing affect. *Applied Linguistics*, 32(2), 149-171.
- *Marshall, K. L., & Bokhorst-Heng, W. D. (2020). Canadian immersion students' investment in French. *Modern Language Journal*, 104(3), 601-617. <https://doi.org/10.1111/modl.12663>
- *McGregor, J., & Fernández, J. (2019). Theorizing qualitative interviews: Two autoethnographic reconstructions. *Modern Language Journal*, 103(1), 227-247. <https://doi.org/10.1111/modl.12541>
- *Miller, E. R., Morgan, B., & Medina, A. L. (2017). Exploring language teacher identity work as ethical self-formation. *Modern Language Journal*, 101(S1), 91-105. <https://doi.org/10.1111/modl.12371>
- *Miranda, J. L., Wells, J. C., & Jenkins, A. (2019). Preparing special education teacher candidates to teach English language learners with disabilities: How well are we doing? *Language Teaching Research*, 23(3), 330-351. <https://doi.org/10.1177/1362168817730665>
- Mirhosseini, S. A. (2020). *Doing qualitative research in language education*. Springer Nature.
- *Mitchell, R., Tracy-Ventura, N., & Huensch, A. (2020). After study abroad: The maintenance of multilingual identity among Anglophone languages graduates. *Modern Language Journal*, 104(2), 327-344. <https://doi.org/10.1111/modl.12636>

- *Moore, A. R. (2019). Interpersonal factors affecting queer second or foreign language learners' identity management in class. *Modern Language Journal*, 103(2), 428-442. <https://doi.org/10.1111/modl.12558>
- Mori, Y. (1999). Epistemological beliefs and language learning beliefs: What do language learners believe about their learning? *Language Learning*, 49(3), 377-415.
- *Mori, J., & Sanuth, K. K. (2018). Navigating between a monolingual utopia and translanguing realities: Experiences of American learners of Yorùbá as an additional language. *Applied Linguistics*, 39(1), 78-98. <https://doi.org/10.1093/applin/amx042>
- *Nagle, C. (2018). Motivation, comprehensibility, and accentedness in L2 Spanish: Investigating motivation as a time-varying predictor of pronunciation development. *Modern Language Journal*, 102(1), 199-217. <https://doi.org/10.1111/modl.12461>
- *Nagle, C., Trofimovich, P., & Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. *Studies in Second Language Acquisition*, 41(4), 647-672. <https://doi.org/10.1017/S0272263119000044>
- *Nakatsukasa, K. (2016). Efficacy of recasts and gestures on the acquisition of locative prepositions. *Studies in Second Language Acquisition*, 38(4), 771-799. <https://doi.org/10.1017/S0272263115000467>
- O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge University Press.
- *Papi, M., & Hiver, P. (2020). Language learning motivation as a complex dynamic system: A global perspective of truth, control, and value. *Modern Language Journal*, 104(1), 209-232. <https://doi.org/10.1111/modl.12624>
- Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4th ed.). Sage.
- *Pawlak, M., Mystkowska-Wiertelak, A., & Bielak, J. (2016). Investigating the nature of classroom willingness to communicate (WTC): A micro-perspective. *Language Teaching Research*, 20(5), 654-671. <https://doi.org/10.1177/1362168815609615>
- *Pekarek Doehler, S., & Berger, E. (2018). L2 interactional competence as increased ability for context-sensitive conduct: A longitudinal study of story-openings. *Applied Linguistics*, 39(4), 555-578. <https://doi.org/10.1093/applin/amw021>
- *Perera, K. (2019). "Cracking a strange veiled code:" Language learning experiences of North Americans in South Asia. *Applied Linguistics*, 40(1), 132-151. <https://doi.org/10.1093/applin/amx013>
- *Pfenninger, S. E. (2020). The dynamic multicausality of age of first bilingual language exposure: Evidence from a longitudinal content and language integrated learning study with dense time serial measurements. *Modern Language Journal*, 104(3), 662-686. <https://doi.org/10.1111/modl.12666>

- Pfenninger, S. E., & Singleton, D. (2021). Is there a best age for second language learning? Evidence from across the lifespan. In R. Woore & E. Macaro (Eds.), *Debates in modern language education* (pp. 52-65). Routledge.
- *Phung, L. (2017). Task preference, affective response, and engagement in L2 use in a US university context. *Language Teaching Research*, 21(6), 751-766. <https://doi.org/10.1177/1362168816683561>
- Plonsky, L., & Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. *Modern Language Journal*, 100, 538-553.
- *Poupore, G. (2016). Measuring group work dynamics and its relation with L2 learners' task motivation and language production. *Language Teaching Research*, 20(6), 719-740. <https://doi.org/10.1177/1362168815606162>
- *Poupore, G. (2018). A complex systems investigation of group work dynamics in L2 interactive tasks. *Modern Language Journal*, 102(2), 350-370. <https://doi.org/10.1111/modl.12467>
- *Prosic-Santovac, D. (2017). Popular video cartoons and associated branded toys in teaching English to very young learners: A case study. *Language Teaching Research*, 21(5), 568-588. <https://doi.org/10.1177/1362168816639758>
- *Qiu, X., & Lo, Y. Y. (2017). Content familiarity, task repetition and Chinese EFL learners' engagement in second language use. *Language Teaching Research*, 21(6), 681-698. <https://doi.org/10.1177/1362168816684368>
- Rallies, S. F., & Rossman, G. B. (2009). Ethics and trustworthiness. In J. Heigham & R. Croker, (Eds.). *Qualitative research in applied linguistics: A practical introduction* (pp. 263-287). Palgrave MacMillan.
- Reid, J. M. (Ed.). (1995). *Learning styles in the ESL/EFL classroom*. Heinle & Heinle.
- *Révész, A., Michel, M., & Lee, M. (2019). Exploring second language writers' pausing and revision behaviors: A mixed-methods study. *Studies in Second Language Acquisition*, 41(3), 605-631. <https://doi.org/10.1017/S027226311900024X>
- Richards, K. (2009). Interviews. In J. Heigham & R. Croker, (Eds.). *Qualitative research in applied linguistics: A practical introduction* (pp. 182-199). Palgrave MacMillan.
- Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning*, 45(2), 283-331.
- Robinson, P. (2012). Individual differences, aptitude complexes, SLA processes, and aptitude test development. In M. Pawlak (Ed.), *New perspectives on individual differences in language learning and teaching* (pp. 57-76). Springer Nature.
- Rothbauer, P. (2008) Triangulation. In L. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 892-894). Sage.
- Ryan, S. (2019). Motivation as an individual difference. In M. Lamb, K. Csizér, A. Henry & S. Ryan (Eds.), *Palgrave Macmillan handbook of motivation for language learning* (pp. 163-182). Palgrave.

- *Sasaki, M., Mizumoto, A., & Murakami, A. (2018). Developmental trajectories in L2 writing strategy use: A self-regulation perspective. *Modern Language Journal*, 102(2), 292-309. <https://doi.org/10.1111/modl.12469>
- *Sato, M. (2017). Interaction mindsets, interactional behaviors, and L2 development: An affective-social-cognitive model. *Language Learning*, 67(2), 249-283. <https://doi.org/10.1111/lang.12214>
- *Sato, T., & McNamara, T. (2019). What counts in second language oral communication ability? The perspective of linguistic laypersons. *Applied Linguistics*, 40(6), 894-916. <https://doi.org/10.1093/applin/amy032>
- *Sauer, L., & Ellis, R. (2019). The social lives of adolescent study abroad learners and their L2 development. *Modern Language Journal*, 103(4), 739-762. <https://doi.org/10.1111/modl.12589>
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129-158.
- Schmidt, R. W. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010, Singapore, December 2-4* (pp. 721-737). Singapore University of Singapore, Centre for Language Studies.
- *Schwartz, M., & Gorbatt, N. (2017). "There is no need for translation: She understands:" Teachers' mediation strategies in a bilingual preschool classroom. *Modern Language Journal*, 101(1), 143-162. <https://doi.org/10.1111/modl.12384>
- Shernoff, D. J. (2013). *Optimal learning environments to promote student engagement*. Springer
- *Spada, N., & Frolich, M. (1995). *COLT observation scheme: Communicative orientation of language teaching coding conventions*. National Centre for English Language Teaching and Research.
- *Stranger-Johannessen, E., & Norton, B. (2017). The African storybook and language teacher identity in digital times. *Modern Language Journal*, 101(S1), 45-60. <https://doi.org/10.1111/modl.12374>
- *Tannenbaum, M., Michalovich, A., & Shohamy, E. (2020). Toward a new multilingual educational policy in Israel: Attitudes and perceptions of teachers and students. *Modern Language Journal*, 104(3), 581-600. <https://doi.org/10.1111/modl.12667>
- *Tavakoli, P., & Hunter, A. M. (2018). Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research*, 22(3), 330-349. <https://doi.org/10.1177/1362168817708462>

- Teimouri, Y. (2017). L2 selves, emotions, and motivated behaviors. *Studies in Second Language Acquisition*, 39(4), 681-709. <https://doi.org/10.1017/S0272263116000243>
- *Teo, T., Hoi, C. K. W., Gao, X., & Lv, L. (2019). What motivates Chinese university students to learn Japanese? Understanding their motivation in terms of "posture." *Modern Language Journal*, 103(1), 327-342. <https://doi.org/10.1111/modl.12546>
- Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative inquiry*, 16(10), 837-851.
- Tracy, S. J. (2019). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact*. John Wiley & Sons.
- *Umino, T., & Benson, P. (2016). Communities of practice in study abroad: A four-year study of an Indonesian student's experience in Japan. *Modern Language Journal*, 100(4), 757-774. <https://doi.org/10.1111/modl.12351>
- Ushioda, E. (2001). Language learning at university: Exploring the role of motivational thinking. In Z. Dörnyei, & R. Schmidt (Eds.), *Motivation and second language acquisition* (Technical Report # 23, pp. 93-125). University of Hawai'i, Second Language Teaching & Curriculum Center.
- *Uzum, B. (2017). Uncovering the layers of foreign language teacher socialization: A qualitative case study of Fulbright language teaching assistants. *Language Teaching Research*, 21(2), 241-257. <https://doi.org/10.1177/1362168815614338>
- *Van Compernelle, R. A. (2019). Constructing a second language sociolinguistic repertoire: A sociocultural usage-based perspective. *Applied Linguistics*, 40(6), 871-893. <https://doi.org/10.1093/applin/amy033>
- Welch, C., & Piekkari, R. (2006). Crossing language boundaries: Qualitative interviewing in international business. *Management International Review*, 46(4), 417-437.
- *White, C. J. (2018). Agency and emotion in narrative accounts of emergent conflict in an L2 classroom. *Applied Linguistics*, 39(4), 579-598. <https://doi.org/10.1093/applin/amw026>
- *Wolff, D., & De Costa, P. I. (2017). Expanding the language teacher identity landscape: An investigation of the emotions and strategies of a NNEST. *Modern Language Journal*, 101(S1), 76-90. <https://doi.org/10.1111/modl.12370>
- Woodrow, L. (2005). The challenge of measuring language learning strategies. *Foreign Language Annals*, 38(1), 90-99.
- *Wu, C. H. (2018). Intercultural citizenship through participation in an international service-learning program: A case study from Taiwan. *Language Teaching Research*, 22(5), 517-531. <https://doi.org/10.1177/1362168817718573>

- *Yang, X., & Tao, X. (2018). Comparing discourse behaviors of a high-rated and a low-rated Chinese EFL teacher: A systemic functional perspective. *Modern Language Journal, 102*(3), 594-610. <https://doi.org/10.1111/modl.12495>
- *Yashima, T., MacIntyre, P. D., & Ikeda, M. (2018). Situated willingness to communicate in an L2: Interplay of individual characteristics and context. *Language Teaching Research, 22*(1), 115-137. <https://doi.org/10.1177/1362168816657851>
- *Yeldham, M., & Gruba, P. (2016). The development of individual learners in an L2 listening strategies course. *Language Teaching Research, 20*(1), 9-34. <https://doi.org/10.1177/1362168814541723>
- *Zheng, X. (2017). Translingual identity as pedagogy: International teaching assistants of English in college composition classrooms. *Modern Language Journal, 101*(S1), 29-44. <https://doi.org/10.1111/modl.12373>
- *Zheng, Y., Lu, X., & Ren, W. (2020). Tracking the evolution of Chinese learners' multilingual motivation through a longitudinal Q methodology. *Modern Language Journal, 104*(4), 781-803. <https://doi.org/10.1111/modl.12672>

APPENDIX

Coding schemes used in the three rounds of analysis

Coding scheme used in first round of data analysis

Title of the article			
Source			
author, journal, volume, issue			
Topic			
main ID topic as defined by authors			
Research approach of the article			
quantitative			
mono-method qualitative			
multi-method qualitative			
mixed methods with qualitative part			
other			

Coding scheme used in the second round of data analysis for the selected qualitative articles

Title of the article	
Type of data collection tools applied	
interview	
focus group interview	
stimulated recall	
think aloud protocol	
reflection	
diary or journal	
narrative	
verbal report	
identity memo	
other type of spoken or written language produced in response to some prompt or question	
observation/recording	
language tasks	
language tests	
questionnaires	
ratings	
scores	
logs	
other measurements	
emails	
books	
blogs	
other naturally occurring texts	
lesson plans	
syllabi	
course books	
other types of teaching materials	

Coding scheme used in the third round of data analysis for articles reporting the use of interviews, observations, or tasks

Title of the article	
Quality control for interviews	
audio or video recording mentioned	
reference to length or corpus size	
description about the content	
interview guide or a sample of it provided in the appendix or in additional materials	
piloting mentioned	
language of interview mentioned	
Quality control for observations	
audio or video recording mentioned	
reference to length or corpus size	
presence of an observation tool (field notes or observation grid) mentioned	
description about the content of the observation tool (if present)	
instrument or a sample of it provided in the appendix or in additional materials (if present)	
piloting mentioned	
Quality control for tasks	
audio or video recording mentioned	
reference to length or corpus size	
description about the content	
task or a sample of it provided in the appendix or in additional materials	
piloting mentioned	
task type receptive or productive	
modality if productive task oral or written	
task type explicitly mentioned	