RESEARCH ARTICLE

Research methods for IDs and TBLT: A substantive and methodological review

Lara Bryfonski [ID], Yunjung (Yunie) Ku [ID] and Alison Mackey

Georgetown University
Corresponding author: Lara Bryfonski; Email: Lara.Bryfonski@georgetown.edu

Abstract

As part of ongoing efforts to characterize the extent to which tasks and interaction-driven language learning are influenced by individual differences (IDs), task-based researchers have thus far examined variables like learners' levels of L2 anxiety, motivation, cognitive creativity, working memory capacity, and aptitude. Building on a tradition of prior syntheses in task-based language teaching (TBLT, e.g., Plonsky & Kim, 2016), we carried out a methodological review of the practices used by researchers who have examined learners' IDs in task-based language learning. We searched journal articles published between 2000 and 2023 and identified 135 unique samples for analysis. Each empirical study was coded for relevant contextual and demographic variables as well as for methodological features related to the investigation of individual differences. We observed that of 30 individual differences investigated in TBLT research over the last two decades, the top five most common were motivation, working memory, L2 proficiency, anxiety, and aptitude. Interesting patterns related to operationalizations, instruments, coding, analyses, and reporting practices. In this paper, we report these results and summarize the most and least common methodological practices, also pointing out gaps and possibilities for future directions. We conclude with recommendations for researchers interested in embarking on empirical investigations of individual differences and TBLT based on best practices.

Introduction

Researchers in the field of second language acquisition (SLA) have long been interested in learners' individual differences (IDs), and the complex role they play in the second language (L2) learning process. For example, Larsen-Freeman and Long's (1991) classic text notes that "it is undeniable that important individual differences between language learners exist" (p. 153). A substantial body of L2 research on IDs has amassed, including motivation (e.g., Dörnyei & Kormos, 2000), L2 anxiety (e.g., Teimouri, Goetze, & Plonsky, 2019), working memory capacity (e.g., Mackey, Adams, Stafford, & Winke, 2010) and aptitude (e.g. Li, 2016; Sparks, 2012), often focusing on how these factors

might moderate L2 development (DeKeyser, 2012; Robinson, 2005). This research includes theoretical, empirical, and meta-analytic studies.

Moving from general SLA work to studies of task-based language teaching (TBLT), an important line of research has focused on how individual differences might help to explain the extent to which learners can benefit from tasks (Awwad & Tavakoli, 2019; Butler & Zeng, 2014; Kim et al., 2015; Sato & McDonough, 2020). Researchers have used a variety of methods and techniques to understand the impact of IDs on task-based interaction and learning, ranging from assessments and interviews to questionnaires and stimulated recalls, amongst others. The current paper presents a methodological review of practices used by researchers studying learner IDs in task-based language learning, with a detailed analysis of what emerged as the top five most frequently investigated IDs in TBLT research to date. We pay particular attention to the instruments, coding, analyses, and reporting practices utilized by researchers in this area, with goals of surveying the domains that have been of greatest interest to researchers, providing empirically-grounded methodological guidance, and highlighting potential avenues for further investigation.

Literature review

The goal of this paper is to examine how IDs are studied within task-based research. Most task-based researchers agree that a task can be broadly defined as an activity with a communicative purpose and a non-linguistic outcome (Ellis, 2018; Long, 2015; Mackey, 2020a). Task-based approaches in the literature vary, including models that follow a pre-task, post-task sequence (Ellis, 2003), those that are based on a task cycle with an element of focus on form (Willis, 1996), and those that follow a sequence of pedagogic tasks approximating real-life target tasks (Long, 2015). Regardless of the approach, task-based researchers and practitioners are interested in how tasks facilitate the kinds of negotiation for meaning and interaction known to support successful SLA (Gass & Mackey, 2006; Mackey, 2020a). Researchers are often also interested in how manipulating specific task-related variables impacts linguistic and non-linguistic outcomes. These variables include increasing the cognitive complexity of the task (e.g., Robinson, 2011a), repeating the task (e.g., Bygate, 2018, Mackey, 1999), or offering planning time (e.g., Bygate & Samuda, 2005). In addition to pedagogic uses, tasks are also used as tools for eliciting oral or written L2 production in empirical SLA investigations (e.g., Housen et al., 2012; Yousefi, 2016).

Research on individual differences

A subset of the research into tasks and second language learning investigates how individual differences among learners might mediate task outcomes and processes. Following Li et al. (2022) and Ortega (2009), individual differences can be broadly categorized into four groups: cognitive (e.g., aptitude), conative (e.g., motivation), affective (e.g., anxiety), and demographic (e.g., age) differences. IDs are generally conceptualized as learner-internal factors, either fixed or changeable, that can affect the process and/or products of second language acquisition and may be mediated by the environment. IDs have been investigated within learners as well as for other interlocutors like teachers (e.g., Bryfonski, 2021) and non-teachers (Gurzynski-Weiss & Plonsky, 2017). However, a few ID variables have garnered sustained attention by second language acquisition researchers for decades: aptitude, working memory, cognitive creativity, motivation, and anxiety.

Aptitude has generally been used to mean cognitive abilities that are posited to be predictive of speed, efficiency, and success in terms of language learning. Carroll's classic (1981) definition claims "an individual's initial state of readiness and capacity for learning a foreign language, and probable facility in doing so given the presence of motivation and opportunity" (p. 86). Aptitude has been a topic of research interest since at least the 1950s (Gass & Mackey, 2012; Skehan, 2015). Aptitude has been measured in a number of ways, and, as our own analysis suggests, researchers tend to believe that there is not one single aptitude factor. For example, some scholars view working memory as a subset of aptitude (e.g., Wen, 2016). Studies that have discussed or measured aptitude and tasks in some way include Yilmaz and Granena (2015), with overviews in Dörnyei and Skehan, (2003), Skehan (2015), and Wen et al. (2017) raising interesting ongoing questions that should be addressed by more research in this area. Aptitude has been the topic of a great deal of interest in the general SLA literature, with theoretical, empirical, and synthetic papers, including a comprehensive and critical synthesis of the methods utilized in studies of aptitude in second language (L2) learning by Li and Zhao (2021).

Working memory capacity is another cognitive area where learners differ. Working memory involves not only storage capacity or what we usually think of when we hear the term "memory" but also processing, which is what is meant by the word "working," in other words, doing something. In an early study in this area, Mackey et al. (2010) looked at the relationship between working memory and output, concluding that individuals with greater working memory capacity produced more modified output in L2 Spanish interaction. Other studies carried out by Kim et al. (2015), Révész (2012), Sagarra (2007), Trofimovich et al. (2007), and Yilmaz and Sağdıç (2019) all point to the fact that working memory capacity is associated with learners' development of the target language and mediated by other learner-external factors such as task complexity and feedback type. In terms of how we assess working memory, most tests originate from research in cognitive psychology, with three that are commonly used in SLA being operation span, counting span, and sentence span (for more information, see Gass et al., 2020).

Differences in learners' levels of cognitive creativity typically involve looking at constructs like originality, elaboration, flexibility, and fluency. Early studies involving cognitive creativity and task performance were carried out by Albert and Kormos (2004, 2011) who demonstrated a relationship between creativity and performance on an L2 narrative task. McDonough et al. (2015) also showed that creativity was associated with the use of questions and coordination in a group problem solving task, and Suzuki et al. (2022) demonstrated a close relationship between creativity and the discourse of speaking tasks. Pipes (2023) provides a helpful overview of research and practice in this area.

A commonly studied conative variable that differs by individual is motivation, which is often seen as how much active, personal involvement in L2 learning there is, as well as how long learners persevere and maintain L2 skills (e.g., Dörnyei, 2009b). One of the earliest studied individual differences in L2 research (e.g., Larsen-Freeman & Long, 1991), motivation has grown dramatically recently with ~277,000 citations in Google Scholar for "motivation in second language acquisition" in the last 10 years, compared with ~74,000 in the 10 years prior. Dörnyei's (2005) highly influential theory of the L2 motivational self-system upended traditional frameworks of motivation and inspired many later studies to investigate motivational thinking as part of learner psychology, concepts of self, and identity. Meta-analytic research (Al-Hoorie, 2018; Yousefi & Mahmoodi, 2022) investigating the L2 motivational self-system has tied motivation to learners' subjective intended effort, underscoring the importance of

motivation as an ID in L2 learning. More recently, Leeming and Harris (2022) have called for using Self-Determination Theory to understand the motivational benefits of tasks within a TBLT framework.

Finally, anxiety, one of the most extensively researched affective factors, has also been shown to vary amongst individual second language learners. What has often been termed "foreign language anxiety" concerns three related performance anxieties: communication apprehension, test anxiety, and fear of negative evaluation (Horwitz et al., 1986). Anxiety can be dynamic, fluctuating throughout tasks that might be associated with changes in linguistic performance (see, for example, Bashori et al., 2022; Papi & Khajavy, 2023). Early research in L2 learning posited optimal levels of anxiety (which introspective measures suggest might be related to tasks and interlocutors) where language learning could be enhanced versus negative levels, which were assumed to be associated with impending anxiety. Baralt and Gurzynski-Weiss (2011) compared learners' state anxiety during task-based interaction in computer-mediated and face-to-face communication, finding learners' reported state anxiety to be comparable across modalities. Current research on anxiety has explored the construct from the perspective of complex dynamic system theory, motivating researchers to delve into the very sources that drive the dynamic nature of anxiety (Papi & Khajavy, 2023). This also encourages practitioners to design pedagogical interventions that may help learners manage anxiety more efficiently.

Syntheses in task-based L2 research

We now turn to our methodological synthesis of current practices in task-based research that has investigated learner IDs. Our general approach follows that used by earlier synthetic research (e.g., Plonsky & Kim, 2016) in that we review substantive and methodological features rather than quantitatively synthesize effect sizes. Prior TBLT meta-analyses have examined the extent to which task-based interaction facilitates the acquisition of grammatical and lexical knowledge by synthesizing effect sizes (Cobb, 2010; Keck et al., 2006; Mackey & Goo, 2007). Mackey and Goo (2007) investigated how different task and design features mediated interaction-driven learning, as well as whether the effects of task-based interaction were durable over time. Ziegler (2016) examined methodological features of task-based interaction research by investigating the context of the interaction focusing on computer-mediated communication (CMC) versus face-to-face (FTF) interaction. She found only a small difference between CMC and FTF interaction, favoring CMC for productive measures, but she cautioned about the stability of the finding due to the lack of delayed posttests in the primary studies.

Other meta-analyses have investigated specific task-based features and variables such as Jackson and Suethanapornkul's (2013) examination of nine studies testing Robinson's Cognition Hypothesis (Robinson, 2001), which resulted in a small but positive effect for accuracy but not fluency when complexity was increased along resource-directing dimensions. Sasayama et al. (2018) subsequently updated the finding that increasing task complexity by manipulating the tense needed to complete tasks ("here and now" versus "there and then") led to greater syntactic complexity whereas manipulating complexity by the number of elements or reasoning demands led to greater lexical complexity (also see Révész, 2009).

While these meta-analyses examined task-based L2 outcomes, other meta-analytic work has examined TBLT from a programmatic perspective. For example,

a meta-analysis by Cobb (2010) built on work investigating task-based interaction (e.g., Mackey & Goo, 2007) by looking at 15 studies of learners performing oral communication tasks, finding differences on outcome measures that examined grammatical knowledge. Another programmatic-based meta-analysis by Bryfonski and McKay (2017) examined 52 studies of longitudinal implementation of TBLT (as defined by primary authors), finding a positive effect for task-based approaches for a variety of learning outcomes as well as positive qualitative stakeholder perceptions.

Finally, there has been methodological work, including syntheses of TBLT research focusing on substantive rather than statistical findings, and methodological choices made by primary authors. Plonsky and Brown (2015), for example, meta-analyzed 18 meta-analyses of corrective feedback (focusing on its role as a key element in interaction-based tasks), finding the domain definitions caused each meta-analysis to draw different conclusions. Plonsky and Kim (2016) examined the substantive and methodological features of task-based learner production research. They analyzed 85 primary studies from 2006 to 2015, concluding, interestingly, that task-based researchers showed a preference for investigations of grammar, vocabulary, accuracy, and interaction with much less focus on pronunciation, pragmatics, and task performance work. In summary, while syntheses of TBLT research to date have reviewed prior studies with a focus on various methodological practices and findings, no studies have yet targeted the role of individual differences in task-based research, which is the goal of the current paper.

Motivation for the study

Given the ongoing interest in both individual differences as they relate to task-based language learning and teaching, and the focus on understanding methodological choices, the current study was guided by the following questions:

1) What are the demographic features of recent task-based research that investigated individual differences?
2) What kinds of individual differences have been investigated in recent task-based research?
3) How have individual differences been operationalized and measured in recent task-based research?
4) What sorts of analyses and reporting practices are most commonly seen in recent task-based research that focuses on individual differences?

Method

To answer these research questions, we carried out a substantive and methodological review, meaning that rather than synthesizing effect sizes (e.g., Cohen's *d*, *r*) from the outcomes of quantitative studies, we systematically examined features of prior research. In doing this, we follow best practices in meta-analytic research recommended by a number of researchers (including, Mackey, 2020b; Norris & Ortega, 2006; Plonsky & Oswald, 2015) and prior methodological synthesis (e.g. Plonsky & Kim, 2016; Plonsky & Oswald, 2015; Plonsky et al., 2020) in the domain of TBLT.

Inclusion and exclusion criteria

To systematically sample prior task-based research that has examined learners' individual differences, we applied the following inclusion and exclusion criteria. The first defining characteristic of included studies was a focus on individual differences in the domain of TBLT.

We took an inclusive perspective on individual difference variables, operationalized from top-down and bottom-up perspectives. Top-down perspectives included the individual differences that commonly appear in texts on tasks and have long histories of being studied in the field (e.g., aptitude and working memory). Bottom-up perspectives included individual differences that emerged from our grounded coding on what types of individual difference variables were included in TBLT studies. Any learner-internal variables that mediated the processes and/or outcomes of second language acquisition were included. Exclusion criteria ruled out studies from non-task-based perspectives, for example, studies that examined individual differences but used linguistic tests like Grammaticality Judgement Tasks (e.g., Yilmaz & Granena, 2019) without tasks being a focus. Also excluded were studies that examined TBLT from non-learner perspectives, such as studies that explored teachers' individual differences (e.g., Bryfonski, 2021), or individual differences that were not examined in light of task-based interventions, implementations, or interactions.

We adopted a similar broad operationalization of both individual differences and TBLT, including, for example, studies that examined TBLT from the perspective of learners' needs, pedagogic tasks approximating target tasks (Long, 2015), task-supported language teaching (as in Ellis et al., 2020), task cycles (as in Willis, 1996) and/or pre-, during- and post-tasks (as in Ellis, 2003; 2018). We included quantitative studies that utilized tasks to examine L2 production or outcome data (e.g. Complexity, Accuracy, Fluency/Complexity, Accuracy, Lexis, and Fluency (CAF/CALF; Bui & Skehan, 2018; Housen et al., 2012; Skehan, 1989) measures, oral or written measures), as well as qualitative studies of learners' perceptions of TBLT and task-based interaction.

Following prior task-based methodological syntheses, we included only published peer-reviewed journal articles, meaning we excluded dissertations, theses, book chapters, conference presentations, and all types of unpublished research.

In statistical meta-analyses, methodologists typically recommend an inclusive approach to avoid publication bias. In other words, only including published studies may lead to positively skewed effect sizes due to the bias for statistically significant findings in academic publishing. However, in the meta-synthesis reported here, we aimed to systematically describe the popular areas, methods, and practices, rather than aggregate statistical effects (see, for example, a similar decision and motivation by Li and Zhao, 2021). So, while book chapters and unpublished work such as theses and doctoral dissertations offer valuable contributions to the field, journal articles tend to have greater visibility and impact in terms of readership, and so we believe they reflect the most current areas of inquiry in this domain, and unpublished, non-referred work can be safely excluded for the purpose of this study. Finally, to limit the scope of our search to only recent, accessible research, we only included studies published between 2000 and 2023, where we expected to see the most growth and interest in IDs in task-based research at the time this study was written. We had to exclude studies that were not available in English as they were not accessible to us. A full list of synthesized studies is available at iris-database.org.

Search techniques

To access the relevant body of literature, four databases were reviewed: Linguistics and Language Behavior Abstracts (LLBA), Google Scholar, Educational Resources

Information Center (ERIC), and Web of Science. We utilized the following terms in various combinations to search these databases: "task-based language teaching," "TBLT," "task supported," "task- based," "language learning," and "individual differences." We then cross-checked our list against articles recently published in eight journals that publish research related to our research questions: *Applied Linguistics, Language Learning, Language Teaching Research, the Modern Language Journal, Studies in Second Language Acquisition, System, TASK Journal, TESOL Quarterly, Language Learning & Technology (LLT), the Annual Review of Applied Linguistics (ARAL),* and *Computer Assisted Language Instruction Consortium (CALICO).* We also examined review articles relevant to our research questions (Chong & Reinders, 2020; Donate, 2022; Ehrman et al., 2003; Li & Zhao, 2021; Nikolov, & Djigunović, 2006; Roberts, 2012; Robinson, 2011b; Smith & González-Lloret, 2021) and cross-checked the reference sections against the results from our database searches.

The total studies retrieved from the databases included 323 possible candidates for inclusion, with 133 studies being ultimately selected based on the inclusion and exclusion criteria discussed above. During the coding process, nine studies that were previously included via the criteria described above were found to be outside the scope of the study (e.g., because they did not use tasks as defined by any of the common standards outlined above) and were excluded. This resulted in a total sample of 133 studies included, contributing 135 unique samples. While we believe our sample paints an accurate and current picture of the domain of ID research in TBLT, of course, we do not believe or claim it is exhaustive. Other search terms, backwards-citation checks, a wider range of journals, and/or larger databases could all have uncovered additional studies. Our lack of time, space, and resources to examine literature not printed in English is also a limitation. Despite these shortcomings, given that we did manage to identify what we view as a substantial sample of included studies, spanning a range of timeframes and journals, we took the sample as sufficiently representative to proceed with the analysis, as shown in Tables 1 and 2.

Table 1 shows that most of the included studies (88.15%) were implemented from 2012 to 2023 while only a few of them (11.85 %) were conducted before 2011.

Coding and analysis

To synthesize the relevant characteristics of the included studies, a coding scheme was developed to extract data from the following key areas: general study characteristics (journal, year, etc.), study context characteristics (country, language, modality, etc.), study participant characteristics (L1s, TLs, learner proficiency levels, etc.), research variables under investigation (IDs, dependent variables, etc.), task and design characteristics (task types, implementations, etc.), ID instrument characteristics (methods), statistical analyses (if applicable), coding methods, and open science practices. These characteristics and coding methods are illustrated in Table 3, with the full coding scheme and data set being available for download on IRIS (iris-database.org). To ensure the coding scheme would effectively obtain the characteristics listed above for our area of interest, the scheme was subjected to pilot and revision coding. The coding scheme was revised and refined before being utilized with the full sample of included studies. We then conducted inter-coder reliability testing. Two coders first discussed the coding scheme together and then independently coded 10 sample studies. The results from those 10 samples were then compared to ensure similar coverage for each coded category. Given the low-inference nature of the

Table 1. Studies of Individual Differences in TBLT from 2000 to 2023

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	total
k	1	0	1	0	0	1	1	1	1	1	4	5	16
%	0.74	0.00	0.74	0.00	0.00	0.74	0.74	0.74	0.74	0.74	2.96	3.70	11.85
Year	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	total
k	9	10	7	11	11	12	7	6	13	9	11	13	119
%	6.67	7.41	5.19	8.15	8.15	8.89	5.19	4.44	9.63	6.67	8.15	9.63	88.15

Table 2. Studies of individual differences in TBLT across journals

Journal	k	%	Journal	k	%
System	13	9.63	Asian Pacific Journal of Second and Foreign Language Education	2	1.48
Language Teaching Research	10	7.41	Frontiers in Psychology	2	1.48
Language Learning	7	5.19	Innovation in Language Learning and Teaching	2	1.48
International Review of Applied Linguistics in Language Teaching (IRAL)	5	3.70	International Journal of Applied Linguistics & English Literature	2	1.48
Studies in Second Language Acquisition	5	3.70	International Journal of English Studies	2	1.48
Language Learning and Technology	4	2.96	Journal of English Language Teaching and Learning	2	1.48
TESOL Quarterly	4	2.96	Journal of Language Teaching and Research	2	1.48
The Journal of Asia TEFL	4	2.96	Studies in Second Language Learning and Teaching	2	1.48
The Language Learning Journal	4	2.96	TASK	2	1.48
Applied Linguistics	3	2.22	The Canadian Journal of Applied Linguistics	2	1.48
Applied Psycholinguistics	2	1.48	Theory and Practice in Language Studies	2	1.48
Asian EFL journal	2	1.48			

Note: Table 2 only includes journals that contributed more than one unique sample. All other journals included in this study contributed only one study to the sample.

coding scheme, the coders achieved 91% agreement after their first meeting (with disagreements in seven categories). To resolve these coding discrepancies, which were mainly in the areas of context of the study (foreign versus second language) and statistical tests used, the ratings from a third coder were used, and the first two coders discussed and agreed upon how to code the disagreed upon data going forward. A second round of interrater reliability was then conducted to ensure reliability of the disagreed upon categories going forward. Two raters coded five additional studies from the sample. Once 100% rating agreement was achieved, the remainder of the studies were split up between two raters.

In terms of analysis, the features listed in Table 3 that were based on categorical coding were analyzed using frequencies and percentages. For continuous data such as *n* sizes, treatment lengths, and number of tests conducted, we examined measures of central tendency and dispersion. For all open-ended items, we collapsed categories where possible and again analyzed them using frequencies and percentages.

Table 3. Coding Scheme Summary

Categories	Characteristics (coded categorically as present/absent)	Characteristics (coded as open-ended)
General study information		Author, year of publication, journal, report type
Study context	Setting (foreign/second/ additional language), classroom-based, lab-based, type of educational setting (elementary, secondary, university, language institute), modality (F2F/CMC)	Country (where the study took place)
Study participants	Learner proficiency levels (beginner/intermediate/ advanced, mixed), heritage language learners	L1s, TLs, *n* size
Research variables		Individual differences examined, dependent and independent variables
Study design and tasks	Design type (quantitative, qualitative, mixed method), pre/post testing, observational, delayed posttest, task modality (oral, written)	Treatment length
ID instruments		Type of instrument used to measure ID variables
Statistical analyses and coding	Frequencies/percentages, correlations, chi-square tests, t-tests, ANOVA, ANCOVA, MANOVA, MANCOVA, post-hocs, factor-analysis, regression, SEM, other non-parametric, grounded theory/thematic coding	Other statistical/non-statistical methods used, number of tests conducted
Open science practices	Availability of tasks/instruments/ data open-access, badges	

*CMC= computer mediated communication; F2F = face to face; SEM = structural equation modeling; TL = target language

Results

RQ1: The Demographic Features of the Recent Task-based Research

Demographics of the sample

The studies we analyzed included 9433 participants with an average *n* size per study of 70 and a range of 6 to 612 participants.

Context

As illustrated in Table 4, our analysis showed that the studies mainly focused on students learning languages in foreign language settings (89.63%), where they had relatively limited access to the target language. Also, the majority of studies were lab-based (62.22%) versus classroom-based studies (37.03%). As documented in studies of trends in applied linguistics research (e.g., Andringa & Godfroid, 2020), the majority of studies took place in university contexts (71.85%), followed by language institutes (17.78%), with a relatively small percentage of studies taking place at the secondary

Table 4. Study context characteristics

Characteristics		k	%*
Settings	Foreign language	121	89.63
	Second language	13	9.63
Context types	Lab-based	84	62.22
	Classroom-based	50	37.04
Educational settings	Elementary	10	7.41
	Secondary	13	9.63
	University	97	71.85
	Language institute	24	17.78
Modality	Face-to-face	116	85.93
	Computer-mediated	19	14.07

*Percentages do not always add up to 100 because some studies met multiple criteria
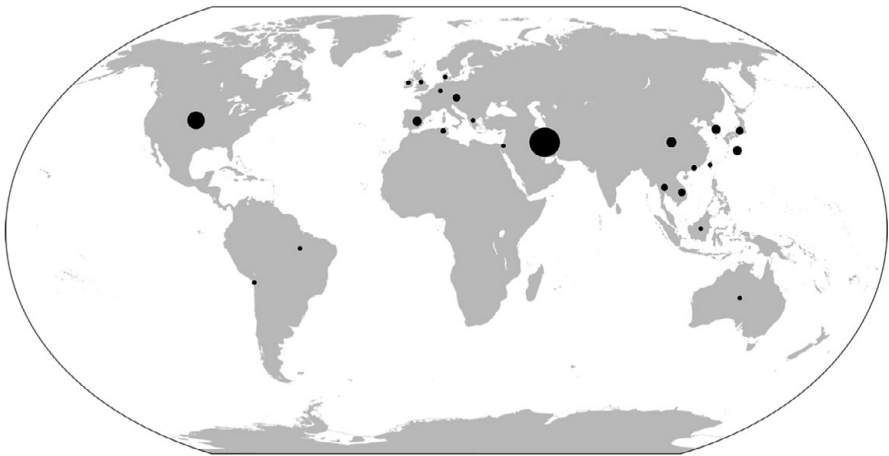
Figure 1. Countries represented by included studies.
Note: The size of the dots represents the number of studies in that region.

(9.63%) or elementary school level (7.41%). Finally, most studies in our sample were conducted in face-to-face modes (85.93%), with the sample also representing a few (*k* = 19) computer-mediated settings.

Participants

Examining the participants within the included studies, we found the majority (43.7%) of participants were rated as intermediate level, non-heritage (94.07%) language learners as illustrated in Table 5. Note that percentages do not add up to 100 because some studies met multiple criteria. The L1 backgrounds of the learners in this sample were varied, with 17.78% of studies examining learners from a mix of L1 backgrounds and a significant portion of the studies (23.70%) not reporting the L1 backgrounds of the learners. This is because we took a strict coding approach to L1 background; for example, when authors described participants as "Chinese learners of English" we did not assume an L1 background of Mandarin (given that, to take just one example, there are hundreds of recognized languages in China, with Mandarin and Cantonese being the two most commonly spoken). For a clearer picture of the range of world regions

Table 5. Participant characteristics

Characteristics		k	%
Learner proficiency levels	Beginner	25	18.52
	Intermediate	59	43.70
	Advanced	6	4.44
	Mixed	33	24.44
Learner backgrounds	Heritage	8	5.93
	Non-heritage	127	94.07
Learner L1s	Arabic	2	1.48
	Cantonese	2	1.48
	Danish	2	1.48
	English	5	3.70
	Farsi	24	17.78
	German	1	0.93
	Hungarian	3	2.78
	Japanese	7	5.19
	Korean	5	4.63
	Mandarin	13	9.63
	Spanish	7	5.19
	Thai	2	1.85
	Vietnamese	6	4.44
	Multiple	24	17.78
	Not clearly specified	32	23.70
Target languages	English	115	85.19
	Spanish	12	8.89
	Korean	3	2.22
	Mandarin Chinese	2	1.48
	German	2	1.48
	French	1	0.74
	Russian	1	0.74

represented by the included studies, we plotted the setting where the study took place in Figure 1.

   In summary, in keeping with previously described trends in applied linguistics research, the majority of the studies we analyzed investigated the learning of English (85.19%) as opposed to other L2s. After English, the only other TLs investigated were Spanish (8.89%), Korean (2.22%), Mandarin (1.48%), German (1.48%), French (0.74%) and Russian (0.74%).

RQ 2: Types of Individual Differences in Recent Task-based Research

To answer Research Question 2: "What kinds of individual differences have been investigated in recent task-based research?" in the included studies, we identified 30 individual differences being studied. We examined both the independent and dependent variables (where applicable) in each included study. For the majority of studies, the independent variables were the individual differences examined in relation to a variety of dependent variables that were typically outcome variables (e.g., anxiety, aptitude, cognitive style, creativity, gender, motivation, personality, prior knowledge, proficiency, and working memory). However, in some cases, IDs also emerged as dependent variables. This is especially the case in motivation research, which often examines the impact of various task manipulations on motivation as an outcome.

The most commonly examined ID was motivation, closely followed by working memory and L2 proficiency. Anxiety, aptitude, gender, prior knowledge, and learner interests were also commonly examined. These findings point to the variety of sub-areas of interest within task-based research, although some of the IDs identified, as illustrated in Table 6, represent overlapping constructs. For example, working memory is often examined as a sub-construct of aptitude. For the purposes of the study reported in this chapter, we coded based on the terms as they were used by primary authors.

RQ 3: Operationalization and Measurement of Individual Differences

To answer Research Question 3, "How have individual differences been operationalized and measured in recent task-based research?", we examined the sorts of instruments used to elicit or measure each of the IDs previously identified to gain insight into how these constructs were operationalized in task-based research. Due to space constraints, this study presents only the five most commonly examined ID variables but the full dataset is available on IRIS (iris-database.org) together with operationalizations and methods for the less commonly examined ID variables.

As noted in relation to Research Question 2 above, the most common ID investigated in the included studies was motivation (30 of 135 studies, or 22.22%). This could be an artifact of time, as motivation research was one of the first individual difference variables to be investigated in L2 research (Larsen-Freeman & Long, 1991). Researchers investigating motivation mainly did so through the use of questionnaires (93.33%) as presented in Table 7. Authors adapted their questionnaires from a variety of pre-existing sources, citing instruments described in Boekaerts (2002), Clément et al. (1994), Gardner (1985), Lam and Law (2007), Martin et al. (1999), Pietri (2015), Pyun et al. (2014), Taguchi et al. (2009), and Troia et al. (2012), amongst others. Gardner's (1985) *Attitudes Motivation Test Battery* and the questionnaire assessing trait-based L2 regulatory focus from Taguchi et al. (2009) were the only materials of this kind to appear in more than one study each. A variety of studies created questionnaires specifically tailored to the study or tasks utilized in the classroom. For example, Torres and Serafini (2016) developed a questionnaire consisting of items related to learners' persistence with the task, interest in the activities, and satisfaction with their performance. Other methods of elicitation included journal entries (Sampson, 2012), thermometer ratings (Azkarai & Kopinska, 2020), and interviews (Ruan et al., 2015).

Six of the motivation studies examined how learners' motivational profiles impacted their L2 production during or after task performance as measured by CALF (e.g., Han & McDonough, 2021). Ten studies examined how various task manipulations or conditions were related to learners' motivation (e.g., Torres & Serafini, 2016). For example, five out of those ten studies examined the relationship between motivation and task complexity, five examined motivation across task types or conditions, and one examined motivation and task repetition. Some of these studies also assessed motivation in conjunction with other IDs such as anxiety, attitudes, task engagement, interest, and proficiency. Studies of how TBLT is mediated by motivation, then, clearly represent rich and interesting areas.

Working memory was the second most commonly investigated ID in task-based research (17.78% of studies, as shown in Table 8). All studies that investigated working memory utilized some form of a memory span task, which can be loosely operationalized as the longest list of items (words, digits, sounds, etc.) a participant can recall. The most commonly used were operation-span tasks (41.67%), where participants complete

Table 6. Research variables

Variables		k	%
Individual differences	Motivation	31	22.96
	Working memory	24	17.78
	L2 proficiency	24	17.78
	Anxiety	16	11.85
	Aptitude	9	6.67
	Willingness to communicate	9	6.67
	Self-efficacy	8	5.93
	Gender	7	5.19
	Prior knowledge	7	5.19
	Engagement	7	5.19
	Interests	6	4.44
	Age	4	2.96
	Cognitive styles	4	2.96
	Personality	4	2.96
	Learning styles	4	2.96
	Creativity	3	2.22
	Self-confidence	3	2.22
	Goal orientation	2	1.48
	Enjoyment	2	1.48
	Self-regulated learning ability	2	1.48
	Emotional intelligence	1	0.74
	Emotions	1	0.74
	Metacognitive strategies	1	0.74
	Heritage identity	1	0.74
	Interaction mindset	1	0.74
	L1 fluency	1	0.74
	Multiple intelligences	1	0.74
	Risk-taking	1	0.74
	Tolerance of ambiguity	1	0.74

math problems, and reading span tasks (29.17%), where participants are asked to read sentences and remember the final word. Studies cited classic reading span tasks by Daneman and Carpenter (1980) and the speaking-span version (Daneman & Green, 1986). Authors also utilized reading span adaptations for other languages such as for Hungarian (Révész, 2012) and Farsi (Shahnazari, 2013). For spatial working memory tasks, authors implemented forward Corsi block-tapping tasks (Zalbidea & Sanz, 2020) or online spatial tasks such as Blockspan and Shapebuilder (Nielson, 2014), both of which ask participants to remember and reproduce flashing or multi-colored shapes in a grid. Several studies note the drawbacks of classic reading-span and listening-span tasks such as Daneman and Carpenter's (1980) for learners who might be asked to complete the tasks in their L2, as justification for using other types of non-language working memory tasks such as spatial memory tasks. The majority of TBLT studies involving working memory (54.17%) investigated the impact of working memory on some dimension of task performance (as measured by CAF/CALF). Five of the included studies investigated the relationship between working memory and corrective feedback during task-based interactions (Goo, 2012; Kim et al., 2015; Lai et al., 2008; Liao & Zhang, 2022; Révész, 2012), and one investigated the production of modified output following corrective feedback (Mackey et al., 2010).

The next most commonly investigated ID in task-based research was L2 proficiency (17.78%). The issue of operationalizing L2 proficiency, namely that it is often not clearly operationalized in applied linguistics research, has been discussed extensively in the

Table 7. Instruments used for investigating motivation in task-based research

Individual differences	Instruments used	k	%
Motivation	Questionnaires	28	93.33
	Journals	1	3.33
	Reflective skits	1	3.33
	Thermometer (adapted from Al Khalil, 2016)	1	3.33
	Written student feedback	1	3.33
	Words written per task	1	3.33

Table 8. Instruments used for investigating working memory in task-based research

Individual differences	Instruments used	k	%
Working memory	Operation span tasks	10	41.67
	Reading span tasks	7	29.17
	Digit span tasks	5	20.83
	Nonword repetition/recognition span	3	12.50
	Speaking span tasks	2	8.33
	Spatial span tasks	2	8.33
	Aural running span tasks	1	4.17
	L1 listening span	1	4.17
	N-back working memory test	1	4.17
	Wechsler (1987) working memory scale	1	4.17

literature (see for example, Bachman and Clark's (1987), early work as well as Malovrh and Benati's (2018) and Park et al.'s (2022) more recent contributions). While it is a frequently used outcome variable in L2 research, we are conceptualizing proficiency as an ID in the current study due to its routine use as an internal mediator of task effects in TBLT research.

We found that studies in task-based research also use a variety of methods to operationalize L2 proficiency (see Table 9). The primary studies we investigated examined the extent to which L2 proficiency mediated L2 outcomes based on a variety of task-related variables such as task complexity (e.g., Awwad & Tavakoli, 2019; Ghahdarijani, 2012; Kim, 2011; Xu & Fan, 2021), pre-task planning (e.g., Bui, 2019) and task type (e.g., oral vs. written, Kim, 2011; or receptive vs. productive, Zareinajad et al., 2015). Studies that investigated L2 proficiency as an ID utilized outcome measures such as CAF (25% of the proficiency studies), listening comprehension (8.33%), interaction/discourse patterns (4.17%; Butler & Zeng, 2014), vocabulary development (8.33%; Kim, 2011), how often learners noticed others' errors (4.17%; Sato & McDonough, 2020), and learners' awareness of L2 pragmalinguistic features (4.17%; Takahashi, 2005). To operationalize L2 proficiency, authors utilized the instruments identified in Table 10. The most common assessment was a standardized TOEFL test (20.83%). Other frequently used assessments included enrollment status in a particular grade (Butler & Zeng, 2014) or class (Kim, 2011) and C-tests (e.g., Dörnyei & Kormos, 2000; Monteiro & Kim, 2020).

The next most commonly examined ID was anxiety (11.85%, see table 10). All of the included studies utilized questionnaires to measure anxiety. One study (Wang et al., 2021) also included semi-structured and stimulated recalls (Gass & Mackey, 2016) to formulate a subsequently developed anxiety questionnaire. Each of the studies utilized or adapted their anxiety questionnaire from a different source, with sources including: the Foreign Language Classroom Anxiety Scale (Horwitz et al., 1986), Abolghasemi's

Table 9. Instruments used for investigating L2 proficiency in task-based research

Individual differences	Instruments used	k	%
L2 Proficiency	TOEFL	5	20.83
	Enrollment status (level, grade)	4	16.67
	C-tests	4	16.67
	Oxford Placement Test	4	16.67
	Institution-created exam	3	12.50
	Elicited imitation task	2	8.33
	General Tests of English Language Proficiency (G TELP)	2	8.33
	Dictation task	1	4.17
	University of Cambridge placement test	1	4.17
	Hong Kong advanced-level public exams (HKALE)	1	4.17
	IELTS	1	4.17
	Informal interview	1	4.17
	National College Entrance Exam (NCEE)	1	4.17
	Oral proficiency test (OPT)	1	4.17
	Teacher evaluation	1	4.17
	Written composition	1	4.17
	X_lex and Y_lex vocabulary test (Meara & Milton, 2003)	1	4.17

Table 10. Instruments used for investigating anxiety in task-based research

Individual differences	Instruments used	k	%
Anxiety	Questionnaires	16	100
	Heart rate	1	6.25
	Semi-structured interviews	1	6.25
	Stimulated recall interviews	1	6.25

Test Anxiety Inventory (Abolghasemi et al.,1996), Brunfaut and Révész (2015), which was adapted from the Foreign Language Listening Anxiety Scale (Elkhafaifi, 2005), Second Language Writing Anxiety Inventory, (Cheng, 2004), MacIntyre, and Gardner (1994), A self-perceived communication competence scale (McCroskey, & McCroskey, 1988), Pyun et al. (2014), Robinson (2001), and Yashima (2002). The Horwitz et al. (1986) scale was identified as the most commonly used instrument to measure anxiety in general L2 research in Teimouri et al.'s (2019) meta-analysis of L2 anxiety and achievement. However, in our sub-set of task-based studies, we found a wider range of approaches being implemented.

In these studies, 37.50% utilized CAF as an outcome measure, while one study utilized listening comprehension assessments (Ghahdarijani, 2012), and one examined the quantity and quality of interactions (Révész, 2011). Six of the studies examined anxiety in conjunction with other IDs such as task motivation (Mahdavirad, 2017; Wang et al., 2021), attitudes (Pyun, 2013), and willingness to communicate (van de Guchte et al., 2022). Researchers also examined how task complexity (56.25%) or task repetition (6.25%) was related to anxiety during task-based interventions.

Aptitude was the fifth most commonly investigated ID in task-based research (6.67%). Many studies that investigated aptitude (44.44% of them) utilized CALF as the outcome measure. The Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959) was the most commonly used method of operationalizing language aptitude in these studies followed by the LLAMA aptitude tests (Kourtali & Révész, 2020; Monteiro

Table 11. Instruments used for investigating aptitude in task-based research

Individual differences	Instruments used	k	%
Aptitude	Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959)	5	55.56
	LLAMA Language Aptitude Tests	2	22.22
	Pimsleur's language aptitude battery	2	22.22
	Hungarian Language Aptitude Test (HUNLAT; Ottó, 2002)	1	11.11
	Oxford Language Aptitude Test	1	11.11

Table 12. Design and task types

Design and task characteristics		k	%
Method type	Quantitative	98	73.30
	Qualitative/interpretivist	4	3.00
	Mixed methods	32	23.70
Test types	Pre/posttests	39	28.89
	Delayed posttests	10	7.41
	CALF measures	43	31.85
Task modalities	Oral	91	67.41
	Written	49	36.30

Table 13. Statistical analyses and coding practices

Test types	k	%
Frequencies/percentages	74	54.81
Correlations	50	37.04
t-tests	39	28.89
ANOVA	34	25.19
Post-hoc tests	27	20.00
Other non-parametric tests	25	18.52
Regression	27	20.00
MANOVA	18	13.33
Grounded theory/thematic coding	21	15.56
Chi-square tests	7	5.19
Factor analysis	9	6.67
ANCOVA	9	6.67
MANCOVA	3	2.22

& Kim, 2020) and Pimsleur's Language Aptitude Battery (Kormos & Trebits, 2012; Li et al., 2019). However, two other aptitude tests were also utilized by task-based researchers in our sample: the Hungarian Language Aptitude test and the Oxford Language Aptitude test (see Table 11).

Researchers investigating aptitude in TBLT did so by examining the relationship between manipulating task complexity and aptitude (44%, all but one manipulated reasoning demands), planning time (22%), or task type (oral vs. written modes, 11%; picture description vs. narrative tasks, 11%).

RQ 4: analyses and Reporting Practices in Recent Task-Based Research

Finally, to answer Research Question 4, "What sorts of analyses and reporting practices are most commonly seen in recent task-based research that focuses on individual differences?", we first looked at the study designs. We found that the majority of the

research was quantitative (72.59%) or mixed methods (23.70%), with the rest being qualitative (3.0%) as illustrated in Table 12. Thirty-nine (28.89%) of the studies were longitudinal, and twenty-eight (20.74%) tracked changes over time using pre/post and/or immediate and/or delayed posttests, although only ten (7.41% of the sample) utilized delayed posttests. On average, the length of treatment in the longitudinal studies was 10 weeks, ranging from one to 40 weeks. More studies utilized oral tasks (67.41%) than written tasks (36.30%); however, both were well represented in the sample. Over a third of the studies utilized some form of CAF measures to examine L2 outcomes.

We next examined the most commonly implemented statistical analyses and coding practices of the included studies. More than a third (37.69%) of the quantitative studies in our sample utilized more than 10 statistical tests per study whereas 7.69% of the included studies ran no statistical tests at all. Around half of the studies (55.38%) ran fewer than 10 statistical tests. Most studies utilized frequencies and percentages (54.81%) followed by correlations (37.04%), *t*-tests (28.89%), and ANOVAs (25.19%) as demonstrated in Table 13. These are slightly different findings for our study than those reported in previous syntheses. In other words, the findings we report here for task-based ID research are not always the same as findings presented in prior methodological syntheses of task research. For example, Plonsky and Kim (2016) found that in task-based learner production studies, ANOVA was the most common test utilized by researchers.

Finally, we examined the sorts of open science practices implemented by authors of included studies. Forty-nine (36.00%) studies made their full tasks available in an appendix or an online repository (IRIS, iris-database.org or The Task Bank, tblt. indiana.edu). Thirty-nine studies (29.00%) made other instruments (such as background questionnaires) available on IRIS. In other words, 74 of 135 studies (54.81%) did not make any tasks or instruments available. Seven studies made their full datasets available, and two studies acknowledged receiving badges for open science. This might be because open science practices have increased in recent years but were seldom practiced in the earlier period for which we collected studies (see Figure 2).
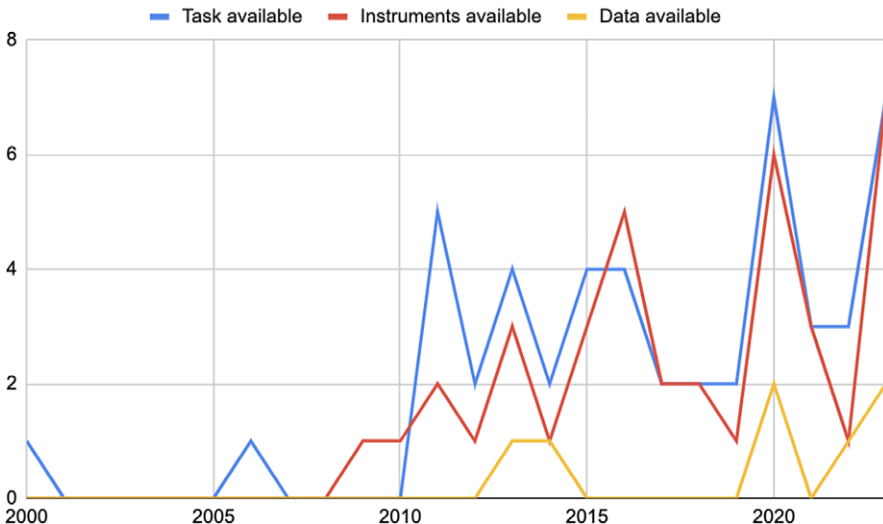

Figure 2. Open Science practices in TBLT ID research over time.

Discussion

Our research provides an overview of the range of IDs investigated in recent, peer reviewed TBLT research along with information about how they are being investigated. We found that this domain of research is growing in popularity, with relatively few articles in this domain published in the early 2000s, up to nearly 10 per year in the 2010s and 15 per year in the 2020s. Our analysis shows that researchers are interested in a diverse array of IDs with motivation, working memory, L2 proficiency, anxiety, and aptitude standing out as the most commonly researched. This finding aligns with interest in L2 research in general where these IDs have robust enough empirical histories to have all been the subjects of other meta-analyses, for example there are prior meta-analyses on motivation (Al-Hoorie, 2018), working memory (Shin, 2020), anxiety (Teimouri et al., 2019), and aptitude (Li, 2016), among others. More than 20 IDs emerged from our analysis, meaning there is ample room for more work in various domains of task-based ID research. Interestingly, ten IDs only appeared in one study each: emotional intelligence, heritage identity, interaction mindset, L1 fluency, multiple intelligences, tolerance of ambiguity, risk-taking, emotions, L2 self-system, and meta-cognitive strategies. This may be due to the fact that some of these IDs can be linked or subsumed into other IDs. For example, L2 risk-taking has been tied to specific domains of personality (Brown, 2000; Pyun et al., 2014). These less commonly investigated IDs point to future potential avenues where task-based ID research might progress.

Our methodological synthesis also uncovered that researchers of the most commonly investigated domains of task-based ID research tend to rely on the same methodological tools. For example, the majority of studies investigating motivation and anxiety relied on questionnaires to operationalize ID variables. This leads us to question whether less commonly implemented tools, for example, those from motivation research, such as journals and written feedback could be triangulated with the more commonly used questionnaires and whether this might lead to a more robust operationalization of the dynamic nature of L2 motivation (e.g., Dörnyei, 2009a).

While Derrick (2016) found that only 58% of L2 studies reported the origins of their instruments, we found for task-based research that authors noted whether they adapted from an existing instrument or developed an instrument in-house for the purposes of the study.

Echoing previous findings in task-based methodological syntheses (Plonsky & Kim, 2016), we found that ID researchers also rely heavily on changes in L2 output based on the CAF/CALF framework (Housen et al., 2012; Skehan 1998a; 1998b; 2009) to operationalize L2 performance and development. Other methods used include assessing listening comprehension, interaction/discourse patterns, vocabulary development, how often learners noticed others' errors, and learners' awareness of L2 pragmalinguistic features.

In terms of the task variables investigated in these studies, our study shows that researchers were mainly interested in investigations of task complexity (27.41%), planning time (12.59%), manipulating task types (11.85%), and corrective feedback (5.93%), among other variables. This range of interests in task-based ID research seems to be representative of domains of interest in TBLT more generally, as evidenced by the recent trends in conferences (Sasayama, 2019), handbooks (Samuda & Bygate, 2008), encyclopedias, and edited collections (Wen et al., 2017) (as noted in a review of recent edited collections by Bryfonski, 2020).

From a methodological standpoint (our fourth research question), only 39 (28.89%) of the studies we investigated were longitudinal, in contrast to 88 (65.19%) that were cross-sectional. Historically, many IDs have been considered to be fixed, unchangeable

characteristics, which may lead researchers to focus on cross-sectional study designs. However, there is also evidence suggesting that IDs like aptitude or working memory might in fact be improvable via training exercises (Bialystok & DePape, 2009; Davidson et al., 2003; Linck et al., 2014). Other studies have found that constructs like motivation or anxiety might be dynamic rather than static, fluctuating by context, including at different times. We are encouraged that for the included longitudinal studies, the average time frame studied was 10 weeks, or slightly less than one academic semester. Many researchers in our field have called for more long-term research (e.g., Long, 2016; Mackey & Goo, 2007). Additionally, the majority of the research we investigated was concentrated in a few contexts, namely, EFL contexts with adult language learners. To move the domain of task-based ID research forward, we believe it is important to recognize the need and value of and support research conducted outside the "WEIRD" (Western, Educated, Industrialized, Rich, and Democratic) contexts traditionally investigated by applied linguists, and social scientists more generally (Andringa & Godfroid, 2020; Henrich et al., 2010). By focusing our investigations mainly on TL (English), the generalizability of findings from these studies of IDs in TBLT is limited.

In terms of our fourth research question, we found that investment in open science practices in the domain of task-based ID research is still developing. Derrick (2016) reported that only 17% of authors in three journals provided instruments in an appendix or in an online repository. We found slightly more (29%) for task-based ID research. Applied linguistics has heralded a push towards open-science practices in recent years, including recognition of open data and materials through badges in major journals (e.g., *Studies in Second Language Acquisition, Annual Review of Applied Linguistics*), repositories for instruments and materials (IRIS, Marsden et al., 2015), repositories for tasks (the Task Bank; Gurzynski-Weiss, 2021), and registered replications and reports (Morgan-Short et al., 2018). Open science practices are an important way to promote scientific equity through the sharing of knowledge, instruments, and findings in freely accessible and permanent repositories. While there is growing excitement around open access in applied linguistics research, practices such as open-access publishing (e.g., Zhu, 2017) or making data freely available have not yet been fully embraced by L2 researchers (and academics more broadly), and this was born out in our findings as well.

Recommendations for Future Research

From a content perspective, the results of this methodological review demonstrate that task-based ID research is expanding beyond the most often studied constructs (motivation, working memory, proficiency). While there is always room for development of studies involving these most commonly researched IDs, we uncovered many other lesser-studied IDs that have the potential to impact TBLT research. To take one example, a few studies have investigated cognitive creativity as an ID (including, for example, Albert & Kormos, 2011; McDonough et al., 2015; Zabihi et al., 2013). IDs like cognitive creativity have the potential to shed light on interesting relationships in how learners approach tasks or task-based interaction, for example investigating how learners' cognitive creativity interacts with their ability to find solutions to task-based problems or utilize learning strategies. However, research in this area has yet to pick up momentum. Less studied IDs, like creativity and emotions, might be profitably combined with other more commonly studied IDs like motivation, (as in Pipes, 2023) to better understand the various ways in which learner IDs mediate outcomes during task-based interactions or interventions.

The task-based ID research that has been conducted so far has relied on a relatively small set of methodological approaches. For example, researchers investigating L2 proficiency could aim to triangulate data from multiple sources in order to present the most accurate, and most transferable, view of participants' developmental levels. This might mean triangulating from standardized test scores in addition to enrollment status and in-house tests or assessments. The results of task-based assessments (e.g., Ellis et al., 2020; Noroozi & Taheri, 2022; Norris et al., 2002) would also be useful to examine in conjunction with other standardized proficiency tests as they are often more representative and better aligned to the kinds of tasks learners complete in task-based interventions (e.g., see Boers et al., 2021).

From a methodological standpoint, we recommend more research focusing on IDs in TBLT from qualitative or mixed methods perspectives. Only four studies (2.96%) included in our methodological synthesis were qualitative, and 32 (23.70%) utilized mixed methods. Again, triangulation of qualitative measures along with quantitative results from questionnaires (the most commonly implemented tool in TBLT ID research according to our findings) such as through semi-structured or stimulated recall interviews, journals, role-plays, classroom discourse, long term case studies, or other qualitative datasets would facilitate our understanding of how learners' individual differences might impact task performance and outcomes. In quantitative studies, we also recommend more longitudinal research that examines changes in L2 outcomes or learners' IDs over time, with a greater focus on longer term effects through the use of delayed posttests or follow-up interviews. Some task-based interventions, such as interactively provided corrective feedback, have been shown to have delayed effects (Lee & Lyster, 2016; Mackey, 1999; Mackey & Goo, 2007; Sheen, 2010). As such, delayed posttests are necessary to observe the contribution of learners' IDs with how durable outcomes are over time.

In the domain of statistical practices, we found that more than a third (37.69%) of the quantitative studies in our sample employed more than 10 parametric statistical tests (and some studies utilized many more). This should be viewed in the light of calls in prior work (e.g., Larsson et al., 2023; Plonsky, 2013; 2015) for researchers to expand their repertoire of statistical practices in quantitative and mixed methods research and prioritize examinations of descriptive statistics, effect sizes, and confidence intervals over running large numbers of null hypothesis statistical tests. In terms of reporting practices too, we recommend authors be explicit about demographic data, including clearly stating the L1s of participants, describing the full context in which the study took place, and including as much descriptive data as possible such that future meta-analytic work can be easily conducted and studies can be replicated if necessary.

Finally, we note that outreach and inclusivity is critical in task-based research. Task-based pedagogy is a worldwide interest and therefore requires a global perspective. We believe an important priority in this area is for research to investigate learners studying languages other than English. While we recognize the global impact of English, our understanding of language learning cannot currently be generalized without the addition of a robust variety of other target languages and in more diverse contexts. Additionally, researchers excited about task-based ID research should consider making their materials such as tasks and data freely accessible in online repositories to aid in replication efforts and to expand the usage of common tools and tasks.

Competing interest. The author(s) declare none.

References

Abolghasemi, A., Asadi Moghaddam, A., Najarian, B., & Shokrkon, H. (1996). Scale reliability for measurement of test anxiety of Ahwaz's guidance school girls. *Journal of Psychology and Educational Sciences of Ahwaz Chamran University*, *3*, 61–74.

Al-Hoorie, A.H. (2018). The L2 motivational self system: A meta-analysis. *Studies in Second Language Learning and Teaching*, *8*, 721–754. https://doi.org/10.14746/ssllt.2018.8.4.2

Al Khalil, M. K. (2016). Insights from measurement of task-related motivation. In A. Mackey, & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS repository of instruments for research into second languages* (pp. 243–262). Routledge.

Andringa, S., & Godfroid, A. (2020). Sampling bias and the problem of generalizability in Applied Linguistics. *Annual Review of Applied Linguistics*, *40*, 134–142. https://doi.org/10.1017/S0267190520000033

Awwad, A., & Tavakoli, P. (2019). Task complexity, language proficiency and working memory: Interaction effects on second language speech performance. *International Review of Applied Linguistics in Language Teaching*, *60*, 169–196, https://doi.org/10.1515/iral-2018-0378

Albert, Á., & Kormos, J. (2004). Creativity and narrative task performance: An exploratory study. *Language Learning*, *54*, 277–310. https://doi.org/10.1111/j.1467-9922.2004.00256.x

Albert, Á., & Kormos, J. (2011). Creativity and narrative task performance: An exploratory study. *Language Learning*, *61*, 73–99. https://doi.org/10.1111/j.1467-9922.2011.00643.x

Azkarai, A., & Kopinska, M. (2020). Young EFL learners and collaborative writing: A study on patterns of interaction, engagement in LREs, and task motivation. *System*, *94*, 1–14. https://doi.org/10.1016/j.system.2020.102338

Bachman, L. F., Clark, J. L. D. (1987) The measurement of foreign/second language proficiency, *The Annals of the American Academy of Political and Social Science*, *490*, 20–33. https://doi.org/10.1177/0002716287490001003

Baralt, M., & Gurzynski-Weiss, L. (2011). Comparing learners' state anxiety during task-based interaction in computer-mediated and face-to-face communication. *Language Teaching Research*, *15*, 201–229. https://doi.org/10.1177/0265532210388717

Bashori, M., Van Hout, R., Strik, H., & Cucchiarini, C. (2022). Web-based language learning and speaking anxiety. *Computer Assisted Language Learning*, *35*, 1058–1089. https://doi.org/10.1080/09588221.2020.1770293

Bialystok, E., & DePape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 565–574. https://doi.org/10.1037/a0012735

Boekaerts, M. (2002). *Motivation to learn: Education practices* (Vol. *10*). International Academy of Education.

Boers, F., Bryfonski, L., Faez, F., & McKay, T. (2021). A call for cautious interpretation of meta-analytic reviews. *Studies in Second Language Acquisition*, *43*, 2–24. https://doi.org/10.1017/S0272263120000327

Brown, H. D. (2000). *Principles of language learning and teaching* (Vol. *4*). Longman.

Brunfaut, T., & Révész, A. (2015). The role of task- and listener-characteristics in second language listening. *TESOL Quarterly*, *49*, 141–168. https://doi.org/10.1002/tesq.168

Bryfonski, L. (2020). Current trends and new developments in task-based language teaching. *ELT Journal*, *74*, 492–511. https://doi.org/10.1093/elt/ccaa043

Bryfonski, L. (2021). From task-based training to task-based instruction: Novice language teachers' experiences and perspectives. *Language Teaching Research*, 1–25. https://doi.org/10.1177/13621688211026570

Bryfonski, L. & McKay, T. H. (2017). TBLT implementation and evaluation: A meta-analysis. *Language Teaching Research*, *23*, 603–632. https://doi.org/10.1177/1362168817744389

Bui, G. (2019). Influence of learners' prior knowledge, L2 proficiency and pre-task planning on L2 lexical complexity. *International Review of Applied Linguistics in Language*, *59*, 543–567. https://doi.org/10.1515/iral-2018-0244

Bui, G., & Skehan, P. (2018). Complexity, accuracy, and fluency. *The TESOL Encyclopedia of English Language Teaching*, 1–7. https://doi.org/10.1002/9781118784235.eelt0046

Butler, Y. G., & Zeng, W. (2014). Young foreign language learners' interactions during task-based paired assessments. *Language Assessment Quarterly*, *11*, 45–75. https://doi.org/10.1080/15434303.2013.869814

Bygate, M. (2018). *Learning language through task repetition*. John Benjamins.

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 37–74). John Benjamins.

Carroll, J. (1981). Twenty-five years of research on foreign language aptitude. In K. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–118), Newbury House.

Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test: MLAT*. The Psychological Corporation.

Cheng, Y. (2004). A measure of second language writing anxiety: Scale development and preliminary validation. *Journal of Second Language Writing*, *13*, 313–335. https://doi.org/10.1016/j.jslw.2004.07.001

Chong, S. W., & Reinders, H. (2020). Technology-mediated task-based language teaching: A qualitative research synthesis. *Language Learning & Technology*, *24*, 70–86. http://hdl.handle.net/10125/44739

Clément, R., Dörnyei, Z., & Noels, K. A. (1994). Motivation, self-confidence, and group cohesion in the foreign language classroom. *Language Learning*, *44*, 417–448. https://doi.org/10.1111/j.1467-1770.1994.tb01113.x

Cobb, M. (2010). *Meta-analysis of the effectiveness of task-based interaction in form-focused instruction of adult learners in foreign and second language teaching*. (Unpublished doctoral dissertation). University of San Francisco, CA.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*, 1–18. https://doi.org/10.1016/0749-596X(86)90018-5

Davidson, R. J., Kabat-Zinn, J., Schumacher, J., Rosenkranz, M., Muller, D., Santorelli, S. F., Urbanowski, F., Harrington, A., Bonus, K., & Sheridan, J. F. (2003). Alterations in brain and immune function produced by mindfulness meditation. *Psychosomatic Medicine*, *65*, 564–570. https://doi.org/10.1097/01.psy.0000077505.67574.e3

DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, *62*, 189–200. https://doi.org/10.1111/j.1467-9922.2012.00712.x

Derrick, D. J. (2016). Instrument reporting practices in second language research. *TESOL Quarterly*, *50*, 132–153. https://doi.org/10.1002/tesq.217

Donate, Á. (2022). Task anxiety, cognition and performance on oral tasks in L2 Spanish. *Journal of Spanish Language Teaching*, *9*, 1–18. https://doi.org/10.1080/23247797.2022.2090661

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah.

Dörnyei, Z. (2009a). Individual differences: Interplay of learner characteristics and learning environment. In N. C. Ellis & D. Larsen-Freeman (Eds.), *Language as a complex adaptive system* (pp. 230–248). Wiley-Blackwell.

Dörnyei, Z. (2009b). The L2 motivational self system. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). Multilingual Matters.

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research*, *4*, 275–300. https://doi.org/10.1177/136216880000400305

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Blackwell.

Elkhafaifi, H. M. (2005). Listening comprehension and anxiety in the Arabic language classroom. *The Modern Language Journal*, *89*, 206–220. https://doi.org/10.1111/j.1540-4781.2005.00275.x

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Ellis, R. (2018). *Reflections on task-based language teaching*. Multilingual Matters.

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press.

Ehrman, M. E., Leaver, B. L., & Oxford, R. L. (2003). A brief overview of individual differences in second language learning. *System*, *31*, 313–330. https://doi.org/10.1016/S0346-251X(03)00045-9

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. Edward Arnold.

Gass, S. M., Behney, J., & Plonsky, L. (2020). *Second language acquisition: An introductory course*. Routledge.

Gass, S. M., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review*, *19*, 3–17. https://doi.org/10.1075/aila.19.03gas

Gass, S. M., & Mackey, A. (2012). *The Routledge handbook of second language acquisition*. Routledge.

Gass, S. M., & Mackey, A. (2016). *Stimulated recall methodology in applied linguistics and L2 research*. Routledge.

Ghahdarijani, M. S. (2012). The impact of task complexity on Iranian EFL learners' listening comprehension across anxiety and proficiency levels. *Theory and Practice in Language Studies*, *2*, 1057–1068. https://doi.org/10.4304/tpls.2.5.1057-1068

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, *34*, 445–474. https://doi.org/10.1017/S0272263112000149

Gurzynski-Weiss, L. (2021). A conversation between task-based researchers, language teachers, and teacher trainers. *TASK*, *1*, 138–149. https://doi.org/10.1075/task.00007.wei

Gurzynski-Weiss, L., & Plonsky, L. (2017). Look who's interacting: A scoping review of research involving non-teacher/non-peer interlocutors. In L. Gurzynski-Weiss (Ed.), *Expanding individual difference research in the interaction approach: Investigating learners, instructors, and other interlocutors* (pp. 305–324). John Benjamins.

Han, Y., & McDonough, K. (2021). Motivation as individual differences and task conditions from a regulatory focus perspective: Their effects on L2 Korean speech performance. *Innovation in Language Learning and Teaching*, *15*, 1–12. https://doi.org/10.1080/17501229.2019.1652614

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83. https://doi.org/10.1017/S0140525X0999152X

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*, 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. John Benjamins.

Jackson, D. O., & Suethanapornkul, S. (2013). The cognition hypothesis: A synthesis and meta-analysis of research on second language task complexity. *Language Learning*, *63*, 330–367. https://doi.org/10.1111/lang.12008

Keck, C. M., Iberri-Shea, G., Tracy-Ventura, N., & Wa-Mbaleka, S. (2006). Investigating the empirical link between task-based interaction and acquisition. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp.91–129). John Benjamins.

Kim, Y. J. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, *61*, 100–140. https://doi.org/10.1111/j.1467-9922.2008.00442.x

Kim, Y. J., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory. *Studies in Second Language Acquisition*, *37*, 549–581. https://doi.org/10.1017/S0272263114000618

Kormos, J., & Trebits, A. (2012). The role of task complexity, modality, and aptitude in narrative task performance. *Language Learning*, *62*, 439–472. https://doi.org/10.1111/j.1467-9922.2012.00695.x

Kourtali, N. E., & Révész, A. (2020). The roles of recasts, task complexity, and aptitude in child second language development. *Language Learning, 70*, 179–218. https://doi.org/10.1111/lang.12374

Lai, C., Fei, F., & Roots, R. (2008). The contingency of recasts and noticing. *CALICO Journal*, *26*(1), 70–90. https://doi.org/10.1558/cj.v26i1.70-90

Lam, S. F., & Law, Y. K. (2007). The roles of instructional practices and motivation in writing performance. *The Journal of Experimental Education*, *75*. 145–164. https://doi.org/10.3200/JEXE.75.2.145-164

Larsen-Freeman, D., & Long, M. (1991). *An introduction to second language acquisition research*. Longman.

Larsson, T., Plonsky, L., Sterling, S., Kytö, M., Yaw, K., & Wood, M. (2023). On the frequency, prevalence, and perceived severity of questionable research practices. *Research Methods in Applied Linguistics*, *2*, 100064. https://doi.org/10.1016/j.rmal.2023.100064

Lee, A. H., & Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition*, *38*, 35–64. https://doi.org/10.1017/S0272263115000194

Leeming, P., & Harris, J. (2022). Self-determination theory and tasks: A motivational framework for TBLT research. *TASK*, *2*, 164–183. https://doi.org/10.1075/task.21024.lee

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, *38*, 801–842. https://doi.org/10.1017/S027226311500042X

Li, S., Ellis, R., & Zhu, Y. (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics*, *40*, 693–722. https://doi.org/10.1017/S0142716418000796

Li, S., Hiver, P., & Papi, M. (2022). *The Routledge handbook of second language acquisition and individual differences*. Routledge.

Li, S., & Zhao, H. (2021). The methodology of the research on language aptitude: A systematic review. *Annual Review of Applied Linguistics*, *41*, 25–54. https://doi.org/10.1017/S0267190520000136

Liao, Y., & Zhang, W. (2022). Corrective feedback, individual differences in working memory, and L2 development. *Frontiers in Psychology*, *13*, 811748. https://doi.org/10.3389/fpsyg.2022.811748

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*, 861–883. https://doi.org/10.3758/s13423-013-0565-2

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley-Blackwell.

Long, M. H. (2016). In defense of tasks and TBLT: Nonissues and real issues. *Annual Review of Applied Linguistics*, *36*, 5–33. https://doi.org/10.1017/S0267190515000057

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. http://doi.org/10.1111/j.1467-1770.1994.tb01103.x

Mackey, A. (1999). Input, interaction, and second language development. *Studies in Second Language Acquisition*, *21*, 557–587. https://doi.org/10.1017/S0272263199004027

Mackey, A. (2020a). *Interaction, feedback and task research in second language learning: Methods and design*. Cambridge University Press.

Mackey, A. (2020b). The Annual Review of Applied Linguistics at 40: Looking back and moving ahead. *Annual Review of Applied Linguistics*, *40*, 1–8. https://doi.org/10.1017/S0267190520000082

Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning*, *60*, 501–533. https://doi.org/10.1111/j.1467-9922.2010.00565.x

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 407–453). Oxford University Press.

Mahdavirad, F. (2017). Affective variables in simple vs. complex tasks: A study of EFL learners' perceptions. *International Journal of English Language & Translation Studies*, *5*, 195–200.

Malovrh, P. A., & Benati, A. G. (2018). *The handbook of advanced proficiency in second language acquisition*. Blackwell Handbooks in Linguistics.

Marsden, E., Mackey, A., & Plonsky, L. (2015). 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.

Martin, M. M., Myers, S. A., & Mottet, T. P. (1999). Students' motives for communication with their instructors and affective and cognitive learning. *Psychological Reports*, *87*, 830–834. https://doi.org/10.2466/pr0.2000.87.3.830

McCroskey, J. C., & McCroskey, L. L. (1988). Self-report as an approach to measuring communication competence. *Communication Research Reports*, *5*, 108–113. https://doi.org/10.1080/08824098809359810

McDonough, K., Crawford, W. J., & Mackey, A. (2015). Creativity and EFL students' language use during a group problem-solving task. *TESOL Quarterly*, *49*, 188–199. https://doi.org/10.1002/tesq.211

Meara, P. & Milton, J. (2003). *X_Lex, the Swansea levels test*. Express.

Monteiro, K., & Kim, Y. (2020). The effect of input characteristics and individual differences on L2 comprehension of authentic and modified listening tasks. *System*, *94*, 1–13. https://doi.org/10.1016/j.system.2020.102336

Morgan-Short, K., Marsden, E., Heil, J., Issa, B., Leow, R. P., Mikhaylova, A., Mikołajczak, S., Moreno, N., Slabakova, R., & Szudarski, P. (2018). Multi-site replication in SLA research: Attention to form during listening and reading comprehension in L2 Spanish. *Language Learning*, *68*, 392–437. https://doi.org/10.1111/lang.12292

Nielson, K. B. (2014). Can planning time compensate for individual differences in working memory capacity? *Language Teaching Research*, *18*, 272–293. https://doi.org/10.1177/1362168813510377

Nikolov, M., & Djigunović, J. M. (2006). Recent research on age, second language acquisition, and early foreign language learning. *Annual Review of Applied Linguistics*, *26*, 234–260. https://doi.org/10.1017/S0267190506000122

Noroozi, M., & Taheri, S. (2022). Task-based language assessment: A compatible approach to assess the efficacy of task-based language teaching vs. present, practice, produce, *Cogent Education*, *9*, 2105775. https://doi.org/10.1080/2331186X.2022.2105775

Norris, J. M., Brown, J. D., Hudson, T. D., & Bonk, W. (2002). Examinee abilities and task difficulty in task-based second language performance assessment. *Language Testing*, *19*, 395–418. https://doi.org/10.1191/0265532202lt237oa

Norris, J. M., & Ortega, L. (2006). *Synthesizing research on language learning and teaching.* John Benjamins.

Ortega, L. (2009). *Understanding second language acquisition*. Hodder Education.

Papi, M., & Khajavy, H. (2023). Second language anxiety: Construct, effects, and sources. *Annual Review of Applied Linguistics*, *43*, 127–139. https://doi.org/10.1017/S0267190523000028

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*, 198–236. https://doi.org/10.1111/lang.12475

Pietri, N. J. M. (2015). The effects of task-based learning on Thai students' skills and motivation. *Asean Journal of Management & Innovation*, *3*, 72–80. https://doi.org/10.14456/ajmi.2015.3

Pipes, A. (2023). *Researching creativity in second language acquisition*. Taylor & Francis.

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. (2015). Statistical power, p values, descriptive statistics, and effect sizes: A "back-to-basics" approach to advancing quantitative methods in L2 research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 23–45). Routledge.

Plonsky, L., & Brown, D. (2015). Domain definition and search techniques in meta-analyses of L2 research (Or why 18 meta-analyses of feedback have different results). *Second Language Research*, *31*, 267–278. https://doi.org/10.1177/0267658314536436

Plonsky, L., & Kim, Y. J. (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*, 583–621. https://doi.org/10.1177/0267658319828413

Plonsky, L., & Oswald, F. L. (2015). Meta-analyzing second language research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 106–128). Routledge.

Pyun, D. O. (2013). Attitudes toward task-based language learning: A study of college Korean language learners. *Foreign Language Annals*, *46*, 108–121. https://doi.org/10.1111/flan.12015

Pyun, D. O., Kim, J. S., Cho, H. Y., & Lee, J. H. (2014). Impact of affective variables on Korean as a foreign language learners' oral achievement. *System*, *47*, 53–63. https://doi.org/10.1016/j.system.2014.09.017

Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition*, *31*, 437–470. https://doi.org/10.1017/S0272263109090366

Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *The Modern Language Journal*, *95*, 162–181. https://doi.org/10.1111/j.1540-4781.2011.01241.x

Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning*, *62*, 93–132. https://doi.org/10.1111/j.1467-9922.2011.00690.x

Roberts, L. (2012). Individual differences in second language sentence processing. *Language Learning*, *62*, 172–188. https://doi.org/10.1111/j.1467-9922.2012.00711.x

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*, 27–57. https://doi.org/10.1093/applin/22.1.27

Robinson, P. (2005). Cognitive complexity and task sequencing: A review studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching*, *43*, 1–33. https://doi.org/10.1515/iral.2005.43.1.1

Robinson, P. (2011a). *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. John Benjamins.

Robinson, P. (2011b). Second language task complexity, the cognition hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 3–37). John Benjamins.

Ruan, Y., Duan, X., & Du, X. Y. (2015). Tasks and learner motivation in learning Chinese as a foreign language. *Language, Culture and Curriculum*, *28*, 170–190. https://doi.org/10.1080/07908318.2015.1032303

Sagarra, N. (2007). Working memory and L2 processing of redundant grammatical forms. In Z. Han (Ed.), *Understanding second language process* (pp. 133–147). Multilingual Matters.

Sampson, R. (2012). The language-learning self, self-enhancement activities, and self-perceptual change. *Language Teaching Research*, *16*, 317–335. https://doi.org/10.1177/1362168812436898

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Springer.

Sasayama, S. (Ed.) (2019). *Proceedings of the TBLT in Asia 2018 conference*. JALT Task-Based Language Teaching Special Interest Group.

Sasayama, S., Malicka, A., & Norris, J. (2018). Cognitive task complexity: A research synthesis and meta-analysis. In Z. Wen & M. J. Ahmadian (Eds.) *Research L2 task performance and pedagogy: In honour of Peter Skehan* (pp. 95–132). John Benjamins.

Sato, M., & McDonough, K. (2020). Predicting L2 learners' noticing of L2 errors: Proficiency, language analytical ability, and interaction mindset. *System*, *93*, 102301. https://doi.org/10.1016/j.system.2020.102301

Shahnazari, M. (2013). The development of a Persian reading span test for the measure of L1 Persian EFL learners' working memory capacity. *Applied Research on English Language*, *2*, 107–116. https://doi.org/10.22108/ARE.2013.15473

Sheen, Y. (2010). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition*, *32*, 203–234. https://doi.org/10.1017/S0272263109990507

Shin, J. (2020). A meta-analysis of the relationship between working memory and second language reading comprehension: Does task type matter? *Applied Psycholinguistics*, *41*, 873–900. https://doi.org/10.1017/S0142716420000272

Skehan, P. (1989). *Individual differences in second language learning*. Routledge.

Skehan, P. (1998a). Task-based instruction. *Annual Review of Applied Linguistics*, *18*, 268–286.

Skehan, P. (1998b). *A cognitive approach to language learning*. Oxford University Press.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*, 510–532. https://doi.org/10.1093/applin/amp047

Skehan, P. (2015) Foreign language aptitude and its relationship with grammar: A critical overview. *Applied Linguistics*, *36*, 367–384. https://doi.org/10.1093/applin/amu072

Sparks, R. (2012). Individual differences in L2 learning and long-term L1-L2 relationships. *Language Learning*, *62*, 5–27. https://doi.org/10.1111/j.1467-9922.2012.00704.x

Smith, B., & González-Lloret, M. (2021). Technology-mediated task-based language teaching: A research agenda. *Language Teaching*, *54*, 518–534. https://doi.org/10.1017/S0261444820000233

Suzuki, S., Yasuda, T., Hanzawa, K., & Kormos, J. (2022). How does creativity affect second language speech production? The moderating role of speaking task type. *TESOL Quarterly*, *56*, 1320–1344. https://doi.org/10.1002/tesq.3104

Taguchi, T., Magid, M. & Papi, M. (2009). The L2 motivational self system among Japanese, Chinese, and Iranian learners of English: A comparative study. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp.66–97). Multilingual Matters.

Takahashi, S. (2005). Pragmalinguistic awareness: Is it related to motivation and proficiency? *Applied Linguistics*, *26*, 90–120. https://doi.org/10.1093/applin/amh040

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, *41*, 363–387. https://doi.org/10.1017/S0272263118000311

Torres, J., & Serafini, E. J. (2016). Microevaluating learners' task-specific motivation in a task-based business Spanish course. *Hispania*, *99*, 289–304. https://doi.org/10.1353/hpn.2016.0055

Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytical ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 171–195). Oxford University Press.

Troia, G. A., Harbaugh, A. G., Shankland, R. K., Wolbers, K. A., & Lawrence, A. M. (2012). Relationships between writing motivation, writing activity, and writing performance: Effects of grade, sex, and ability. *Reading & Writing*, *26*, 17–44. https://doi.org/10.1007/s11145-012-9379-2

van de Guchte, M., van Batenburg, E., & van Weijen, D. (2022). Enhancing target language output through synchronous online learner-learner interaction:The impact of audio-, video-, and text-chat interaction on learner output and affect. *Journal on Task-Based Language Teaching and Learning*, *2*, 218–247. https://doi.org/10.1075/task.21003.guc

Wang, Q., East, M., & Li, S. (2021). Measuring Chinese EFL learners' motivation and anxiety when completing a video narration task: Initial steps in designing two questionnaires. *System*, *100*, 102559. https://doi.org/10.1016/j.system.2021.102559

Wechsler, D. (1987). Wechsler memory scale-revised manual. *Psychological Cooperation Inc.*

Wen, Z. E. (2016). *Working memory and second language learning*. Multilingual matters.

Wen, Z., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, *50*, 1–31. https://doi.org/10.1017/S0261444816000276

Willis, J. (1996). *A framework for task-based learning*. Longman.

Yashima, T. (2002). Willingness to communicate in a second language: The Japanese EFL context. *The Modern Language Journal*, *86*, 54–66. https://doi.org/10.1111/1540-4781.00136

Yilmaz, Y., & Granena, G. (2015). The role of cognitive aptitudes for explicit language learning in the relative effects of explicit and implicit feedback. *Bilingualism: Language and Cognition*, *19*, 147–161. https://doi.org/10.1017/S136672891400090X

Yilmaz. Y., & Granena, G. (2019). Cognitive individual differences as predictors of improvement and awareness under implicit and explicit feedback conditions. *Modern Language Journal*, *103*, 686–702. https://doi.org/10.1111/modl.12587

Yilmaz, Y., & Sağdıç, A. (2019). The interaction between inhibitory control and corrective feedback timing. *International Journal of Applied Linguistics*, *170*, 204–227. https://doi.org/10.1075/itl.19010.yil

Yousefi, M. A. (2016). The influence of affective variables on the complexity, accuracy, and fluency in L2 oral production: The contribution of task repetition. *Journal of English Language Teaching and Learning*, *17*, 25–48.

Yousefi, M., & Mahmoodi, M. H. (2022). The L2 motivational self-system: A meta-analysis approach. *International Journal of Applied Linguistics*, *32*, 274–294. https://doi.org/10.1111/ijal.12416

Xu, J., & Fan, Y. (2021). Task complexity, L2 proficiency and EFL learners' L1 use in task-based peer interaction. *Language Teaching Research*, *28*, 1–20. https://doi.org/10.1177/13621688211004633

Zabihi, R., Rezazadeh, M., & Ansari, D. N. (2013). Creativity and learners' performance on argumentative and narrative written tasks. *The Journal of Asia TEFL*, *10*, 69–93.

Zalbidea, J., & Sanz, C. (2020). Does learner cognition count on modality? Working memory and L2 morphosyntactic achievement across oral and written tasks. *Applied Psycholinguistics*, *41*, 1171–1196. https://doi.org/10.1017/S0142716420000442

Zareinajad, M., Rezaei, M., & Shokrpour, N. (2015). The effects of receptive and productive task-based listening activities on the listening ability of Iranian EFL learners at different proficiency levels. *Pertanika Journal Social Sciences & Humanities*, *23*, 537–552.

Zhu, Y. (2017). Who support open access publishing? Gender, discipline, seniority and other factors associated with academics' OA practice. *Scientometrics*, *111*, 557–579. https://doi.org/10.1007/s11192-017-2316-z

Ziegler, N. (2016). Taking technology to task: Technology-mediated TBLT, performance, and production. *Annual Review of Applied Linguistics*, *36*, 136–163. https://doi.org/10.1017/S0267190516000039