Language specific dual-task effects after stroke:

A systematic review

Christos Salis[1*], Rawand Jarrar[2], and Laura L. Murray[3]

[1]Speech & Language Sciences
Newcastle University
Newcastle upon Tyne
United Kingdom

* Corresponding author
Email: christos.salis@ncl.ac.uk

[2]Population Health Sciences Institute
Newcastle University
Newcastle upon Tyne
United Kingdom

[3]School of Communication Sciences & Disorders
Faculty of Health Sciences
Western University
London, Ontario
Canada

Author contribution statement: All authors contributed equally in this research report.

Ethics statement: This research report is exempt from ethical review because it utilised secondary data that are in the public domain

Abstract

Purpose: The dual-task paradigm has been frequently used to examine stroke-related deficits because it samples behavioural performance under conditions of distraction similar to functioning in real-life environments. This original systematic review synthesizes of studies that examined dual-task effects involving spoken language production in adults affected by stroke, including transient ischaemic attack (TIA) and post-stroke aphasia.

Method: Five databases were searched (inception to March 2022) for eligible peer-reviewed articles. The 21 included studies reported a total of 561 stroke participants. Thirteen studies focused on single word production, e.g., word fluency, and eight on discourse production, e.g., storytelling. Most studies included participants who had suffered a major stroke. Six studies focused on aphasia, whereas no study focused on TIA. A meta-analysis was not appropriate because of the heterogeneity of outcome measures.

Results: Some studies of single word production found dual-task language effects whereas others did not. This finding was compounded by the lack of appropriate control participants. Most single word and discourse studies utilised a motoric task as the dual-task condition. Our certainty (or confidence) assessment was based on a methodological appraisal of each study and information about reliability/fidelity. As only 10 of the 21 studies included appropriate control groups and limited reliability/fidelity information, the certainty of the findings may be described as weak.

Conclusions: Unlike single word studies, nearly all studies of discourse showed dual-task costs in some variables. Future studies should include appropriate neurotypical controls and more detailed and consistent descriptions of motor speech, language and broader cognitive skills.

249 words

Introduction

Our ability to produce language for effective communication and participation in everyday life requires the seamless and timely integration of all aspects of cognition. Workplace, recreational, and domestic settings where people interact with others are rarely free of distraction or multi-tasking. Consequently, being able to resist distraction, focus on what one wishes to say and simultaneously carry on doing other activities require a great deal of cognitive resources, which are often negatively affected in stroke (Stolwyk et al., 2014). A plethora of studies from diverse methodological paradigms have documented the psychosocial impact and nature of the cognitive and communication deficits of stroke (Tang et al., 2018; Tatemichi et al., 1994), as well as more recently, transient ischaemic attacks (TIAs) (Jokinen et al., 2015). Despite being positively responsive to both spontaneous and treatment-induced recovery (Brady et al., 2016; Loetscher et al., 2019), such impairments are most commonly chronic and adversely affect the well-being and meaningful, social and vocational participation of affected individuals, even when impairment severity is very mild (Harmon et al., 2019; Kontou et al., 2020) or even subjective (van Rooij et al., 2017).

The range of names that have been used to describe cognitive impairments after stroke varies. For example, the term 'post-stroke cognitive impairment' has been used (Lo et al., 2019), whereas other times the deficit is labelled after the affected cognitive domain such as attention or working memory impairment (Spaccavento et al., 2019). In the case of impaired language after left hemisphere stroke, the term aphasia is used and preferred (Lazar & Boehme, 2017). While post-stroke 'aphasia' is a prototypical communication impairment, cognitive impairments that also affect communication, not described as aphasia, are frequent after right hemisphere stroke (Lehman Blake, 2018). The term 'apragmatism' has also been suggested as an appropriate label for communication impairments after right hemisphere stroke (Minga et al., 2022).

When it comes to identification of post-stroke cognitive impairments including aphasia, much of the knowledge stems from the classical neuropsychological testing paradigm. This paradigm utilises pen-and-paper or analogous computerised tasks, which either cut across cognitive domains, e.g., MoCA (Nasreddine et al., 2005), RBANS (Randolph, 2012), or scrutinise a specific domain, e.g., memory via the Wechsler Memory Scale [WMS] (Wechsler, 2009) or language via the Western Aphasia Battery-Revised [WAB-R], (Kertesz, 2007). However, issues of diagnostic accuracy have been raised (Blake, 2002; Murray & Coppens, 2021), especially in terms of these tests' ability to detect residual or latent deficits (DeDe & Salis, 2020; Hillis & Tippett, 2014; Kemper et al., 2006). Assessing cognition is important in stroke patients of mild severity because even subtle deficits can influence treatment options and discharge recommendations, and consequently, long-term outcomes (Jaywant et al., 2019). Another shortcoming of neuropsychological tests is their limited ecological validity, since their contents do not resemble spontaneous real-life situations and activities (Chaytor & Schmitter-Edgecombe, 2003).

In contrast to the neuropsychological paradigm, a defining feature of the dual-task paradigm is that it samples behavioural performance under conditions of distraction. In the simplest version (Strobach et al., 2018), dual-task performance, e.g., walking while talking is contrasted with single-task performance, e.g., walking, with only one task (performance on which is of primary interest) being presented in the single-task condition. Thus, participants either perform one task in isolation or two tasks simultaneously in single- and dual-task conditions, respectively. In dual-task conditions, there is interference between the two tasks (Pashler, 1994). Such conditions are not a typical feature of most neuropsychological tests, although there are rare exceptions, such as the Test of Everyday Attention (TEA) (Robertson et al., 1996). The dual-task paradigm has been popular in investigating stroke-related deficits and their remediation (especially motoric physical) presumably because it affords a potential glimpse of a person's functioning in real-life environments and their ability to multi-task (Plummer et al., 2013; Strobach et al., 2018). However, the dual-task paradigm is not uniform. Al-Yahya and colleagues (Al-Yahya et al., 2011) provided a dual-task classification which includes five task categories: reaction time, discrimination and decision-making, mental tracking, working memory, and verbal fluency. Nevertheless, a common feature in dual-task studies is how performance is compared between single- versus dual-task performance, with the relative change in performance between dual- and single-tasking described as the dual-task effect (Plummer & Eskes, 2015). When the dual-task effect is negative from zero, it is described as a dual-task cost which suggests decrement of performance in the variable of interest. In the literature, there are also studies that do not compare performance between a single- versus a dual-task but instead contrast performance between simpler versus more complex dual-tasks. For example, word fluency has been compared while participants were walking forward versus while they were walking backwards (Yang et al., 2016).

The cognitive domain that is implicated in the dual-task paradigm is attention and its various subcomponents. Attention is typically conceived as a hierarchically organised construct, the precise architecture of which varies among researchers, although there some commonalities. For example, Oberauer (2019), echoing to some extent older views (Shiffrin & Schneider, 1977), conceptualised attention broadly across automatic versus controlled processes. Automatic attentional processes are neither resource limited nor voluntary whereas controlled ones are resource-intensive and capacity-limited. Controlled processes are also fuelled by a central processing attention capacity, which is amodal (Hula & McNeil, 2008). In other words, tasks of all types, e.g., motor and cognitive, draw from, at least in part, one central reservoir of attentional capacity. Within this central capacity further theoretical ramifications have been proposed (Pashler, 1994). In more neuropsychologically-informed models, attention is more complex. For example, Sohlberg and Mateer (Sohlberg & Mateer, 2001) proposed five levels: focused, sustained, selective, alternating, and divided attention. Divided attention refers to the ability to process more than one piece of information at a time or to engage in more than one activity simultaneously (Cristofori & Levine, 2015; Kahneman, 1973), akin to the demands imposed in dual-task paradigms. More recently, Sohlberg and Mateer (2010) subsumed divided attention within alternating attention to acknowledge that multi-tasking situations require rapidly alternating or shifting attention. Under this view, the demands in dual-task paradigms increase the need for rapidly alternating or shifting the focus of attention across the two or more tasks a person is called to perform. Despite the diverse theoretical ramifications that are abundant in the typical and neuropsychological literature, attention difficulties are commonplace after stroke (Kemper et al., 2006; Murray, 2012; Spaccavento et al., 2019).

The guiding motivation in the present study was to synthesize for the first time the evidence pertaining to deficits of attention as measured by dual-tasks with respect to language production in people affected by stroke. As discussed earlier, communication impairments after stroke can emanate from aphasia after left hemisphere stroke or from other types of language and communication problems after TIAs or stroke affecting other parts of the brain, e.g., right hemisphere stroke or cerebellar stroke. Thus, the present systematic review focuses on dual-task language production studies in people affected by stroke, minor strokes such as TIA, and post-stroke aphasia. In a dual-task language study, a person carries out a language task while concurrently performing another task, e.g., telling a story while walking, and the single, comparator task is a similar language task, e.g., telling a story while seated. Although there have been recent systematic reviews on dual-task performance after stroke, much of this work has focused on observational or treatment studies that examined dual-task effects on primarily physical abilities post-stroke (He et al., 2018; Hofheinz et al., 2016; Lee & Jung, 2016; Plummer & Iyigün, 2018; Tsang et al., 2022; Won et al., 2020; Zhou et al., 2021). Two other systematic reviews are relevant to mention: i) Deblock-Bellamy et al. (Deblock-Bellamy et al., 2020) centred on dual-task effects on post-stroke cognition but with an emphasis on cognitive-motor aspects, with only three of their included studies focusing on language production; and ii) Poulin et al. (Poulin et al., 2012) focused on dual-task training of stroke patients and the effect of such training on executive function outcome measures, but none of their included studies examined language production. Accordingly, the present systematic review uniquely explores dual-task effects on language production abilities in people affected by stroke across severities, including TIAs. There is growing evidence that TIAs can result in chronic cognitive impairments (Nicolas et al., 2020; Turner et al., 2019; van Rooij et al., 2017). Additionally, our searches also focused on post-stroke aphasia as this literature domain is different from the broader stroke literature. To our knowledge no previous systematic review (published or in progress; the authors searched PROSPERO and the Open Science Framework for related reviews) has focused exclusively on dual-task language production studies in stroke whereby language performance was compared in single- and dual-task conditions. This is another unique feature of the current study. We focus on production of spoken language as opposed to receptive or written language (reading, writing) to manage the rather large volume of stroke dual-task literature.

The aims of the present systematic review are as follows:

1) To identify and describe dual-task studies that focused on language production in people affected by stroke, minor strokes (including TIAs), and post-stroke aphasia, including their findings and participant and methodological characteristics.
2) To critically appraise the current knowledge base, identify gaps and recommend areas for future research.

Method

The method followed several of the PRISMA 2020 checklist items (Page et al., 2021). However, because of the heterogeneity of study designs and differences in outcome measures among studies, most checklist items could not be addressed. For the same reasons, it was not possible to conduct a meta-analysis. However, we followed the SWiM guidelines (Campbell et al., 2020). Completed PRISMA (including abstract) and SWiM checklists can be found in the supplementary materials (S1-S3). This systematic review was not pre-registered. A protocol was not prepared. The review was financially supported by Newcastle University who had no other part in the review. The authors had no competing interests.

Information sources

After consultation with an evidence synthesis specialist, we searched the following databases: EMBASE (via Ovid), Linguistics and Language Behavior Abstracts (LLBA; via ProQuest), Medline (via Web of Science), PsycINFO (via Ovid), and Web of Science Core Collection (via Web of Science). These databases were searched twice: i) from each database's inception to December 2020; and ii) January 2021 to 4 March 2022. Inception dates were as follows: EMBASE – 1974; LLBA – 1970; Medline – 1950; PsycINFO – 1806; Web of Science – 1950. Depending on the database, our searches were limited to "abstract" or "title and abstract" search fields. The search terms are detailed in Table 1. Terms within each of these three categories were combined with OR, followed by AND across categories to derive a final list of records. Full search details from all five databases with information about additional filters and limits can be found in the supplementary materials (S4). The electronic searches were supplemented by hand searches of the reference lists of included articles. Relatedly, hand searches included searching the reference lists of five reviews or systematic review papers that focused on dual-task studies (Deblock-Bellamy et al., 2020; Ghai et al., 2017; Plummer et al., 2013; Wang et al., 2015; Woollacott & Shumway-Cook, 2002).

Table 1 about here

Selection process and eligibility criteria

The electronic records (title and abstract) were exported to EndNote online. Reviewer pairs (authors) worked independently and screened the title and abstract of each record in EndNote online. All authors were reviewers and were apportioned specific sets of the electronic records in pairs. For an article to be included in the review it had to: i) be peer-reviewed; ii) be published in English; iii) include in its majority (> 80%) adults (18 or older) who had suffered a stroke of any severity (including TIA); and iv) include spoken language production data elicited under single- and dual-task conditions. Since language production is required in tasks that assess cognitive domains beyond language, e.g., attention, memory, executive functions, we only included studies that reported primarily language outcomes according to accepted levels of linguistic description (word, sentence, discourse). For example, immediate word recall or word learning tasks while verbal, do not primarily assess language but rather verbal aspects of short- and/or long-term memory (Al-Yahya et al., 2011). Similarly, executive cognitive control tasks, e.g., subtraction from 100 serially by seven, that involve number production, which is a special type of vocabulary, are not considered primarily language tasks (Folstein et al., 1975; Nasreddine et al., 2005). However, word fluency tasks, which require the production of words from a specific semantic, e.g., animals or supermarket items, or phonemic category, e.g., words beginning with "s" or "m", are considered executive function tasks with language as a critical component (Whiteside et al., 2016). Therefore, studies of word fluency were included. In studies that compared performance from more than two tasks (no distraction vs. focused attention vs. divided attention) we extracted information only between the no distraction and the most challenging condition (i.e., divided attention condition). This decision affected two studies (Murray, 2000; Murray et al., 1998). In studies that compared simpler versus more complex language dual-tasks, we extracted information from the simplest and most complex dual-task. This decision affected three studies (Tsang et al., 2019; Tsang & Pang, 2020; Yang et al., 2016). In terms of study design, observational, treatment, single-case experimental studies or case-series experimental studies were included. Studies were excluded if they did not meet one or more of these criteria.

At the end of the title and abstract screening, each reviewer pair generated a mutually agreed number of records for full text screening after consensus discussions. In the full text selection stage, the reviewer pairs, who had worked independently and after consensus discussions, reached the decision as to whether an article met the inclusion criteria. Reasons for exclusion were recorded. When disagreements occurred within the reviewer pairs, a third reviewer who had not been involved in the screening, reviewed the disputed record. Similar processes were followed for records identified through hand searches. All three authors were involved in the selection process.

Data collection, synthesis, critical appraisal

From each included report, one reviewer pairs extracted data independently. If a required piece of information was not included in the report, the missing information was recorded as "not reported". We also assumed that it had not been collected. Tabulated data items were extracted to address aim 1 of the review. Accordingly, the following relevant data were extracted in tables that presented: i) participant information (number and subgroups, demographics, cognitive profile, information about motor speech disorders, information about aphasia or language ability). ii) stroke information (type [i.e., haemorrhagic or ischaemic], lesion site, stroke severity, time post-onset); iii) dual-task information including information about task prioritisation, i.e., whether participants were given specific instructions to dedicate more effort to one of the two tasks they were carrying out when attempting the dual-task, and reliability/fidelity analyses to which the language output data were subjected. iv) findings (single-dual task comparisons, dual-task effect findings). Extraction of information focused only on linguistic variables as opposed to variables from other cognitive or motor abilities. In studies where more than one measurement data point was reported, e.g., pre/post treatment, longitudinal, we extracted data only from the first measurement. In studies where only descriptive statistics (means, SDs) were reported or the results were not clear (i.e., Harley et al., 2006), we conducted inferential statistical analyses using either reported raw data or obtained 95% confidence intervals based on reported means, SDs, and sample size. In studies where subgroup analyses were carried out, the information we extracted was from all stroke participants rather than subgroups.

The above data items formed the basis for a qualitative synthesis as well as a way of examining heterogeneity. The synthesis was primarily qualitative and was based on two broad levels of linguistic description: i) studies that focused on single word production and ii) studies that focused on discourse production. Because of the heterogeneity among the spoken language output variables and the tasks by which they were elicited, a meta-analysis was not deemed appropriate.

To address aim 2, a methodological appraisal of each study was carried out using a tool (see supplementary materials – S5) based on previous work (Murray et al., 2018; Salis et al., n.d.) informed by the Guidance for Undertaking Reviews in Healthcare (Dignen, 2009), and STARD checklists (Bossuyt et al., 2003). This tool focuses on four categories: study design, demographic variables, stroke variables, and cognitive variables. Quality ratings of high, moderate, or low were assigned to each quality category as well as an overall study rating. For a study to receive an overall high rating, three of the four appraisal categories had to attain high ratings with no category obtaining a low rating. A study with a moderate rating could also not have any low rating. This methodological appraisal also involved two reviewer pairs who worked independently and then discussed their ratings to resolve any disagreements. Our certainty (or confidence) assessment (as per PRISMA, 2020 and SWiM 2020) was based on this methodological appraisal of the study and information about reliability/fidelity of language outcome measures, e.g., inter- and intra-rater reliability for word and utterance counts in discourse studies.

Results

The results of the record selection process are shown in Fig. 1, with 21 studies that were included. The results of the data synthesis are reported according to the two study categories, that is, single word and discourse production studies. We found no studies in either category that focused on TIA.

Fig. 1 about here

Single word production studies

Study characteristics and the individual results of the 13 studies that reported single word production data are shown in Tables 2-5. Table 6 shows these studies' critical appraisal ratings.

Regarding participant information (Table 2), 420 stroke participants in total were reported in these studies. Six of the 13 studies included a neurotypical control group. Reported demographic variables included age, education, and sex in eight studies. Five studies reported less sample demographic information. Information about participants' cognitive profiles varied greatly in terms of the number and range of tests used and cognitive abilities assessed. One study (Bhatt et al., 2016) reported no information about the cognitive abilities of participants. Six studies used a screening cognitive test, i.e., MMSE or MoCA. Only one study Bruehl et al. (2021) reported information about presence of a motor speech disorder (apraxia of speech in this case). Participants' language abilities were formally assessed using language or aphasia tests in four studies (Bruehl et al., 2021; Feld & Plummer, 2021; Laganaro et al., 2019; Murray, 2000). Of these studies, Feld and Plummer did not specifically focus on people with aphasia, whereas the other studies were aphasia-specific. In the remaining nine studies, the language abilities of participants were either not reported (Tsang et al., 2019; Tsang & Pang, 2020) or reported as adequate or sufficient without mention of how adequacy/sufficiency was established, e.g., Cockburn et al. (2003), Yang et al. (2016). To summarise, in single word studies, 35 (of 420) participants presented with aphasia (Bruehl et al., 2021; Laganaro et al., 2019; Murray, 2000).

Table 2 about here

Stroke-related information is shown in Table 3. Stroke type (ischaemic or haemorrhagic) was reported in five of the 13 studies, with most participants having suffered ischaemic strokes. Lesion site was not reported in five studies. As for stroke site, i.e., left/right hemisphere, there was a similar number of participants (55 left vs. 59 right) across studies. In Feld and Plummer (2021), 19 participants had suffered a subcortical stroke; this was also the only single word production study in which initial stroke severity (NIHSS scale) was reported. For time post-onset there was variability, with most studies (9/13 studies) including participants both in early (< 3 months) as well as chronic time post-onset (> 3 months), e.g., Cockburn et al. (2000), Haggard et al. (2000). Participants in Feld and Plummer (2021) were in the subacute stage (median = 14 days).

Table 3 about here

Table 4 shows information about task prioritisation and reliability/fidelity. In five studies task prioritisation information was absent (Bruehl et al., 2021; Harley, 2006; Kim, 2021; Laganaro et al., 2019; Tsang & Pang, 2020). In five studies the instructions were about performing both tasks equally well (Bhatt et al., 2016; Cockburn et al., 2003; Feld & Plummer, 2021; Haggard et al., 2000; Tsang et al., 2019; Yang et al., 2016). Participants were instructed to prioritise one of the two tasks in two studies and both involved prioritisation of the non-language task (Lee et al., 2021; Murray, 2000). As for reliability/fidelity, nine studies did not report any such information. In four studies information about reliability/fidelity was detailed (Laganaro et al., 2019; Murray, 2000; Tsang et al., 2019; Yang et al., 2016).

Table 4 about here

Table 5 shows single- versus dual-task comparisons and results of dual-task effects. Recall that in three studies we considered the more complex dual-task as the more demanding task (Tsang et al., 2019; Tsang & Pang, 2020; Yang et al., 2016). In terms of language task types, eight studies used a timed word fluency task, with eight of these studies involving semantic fluency, e.g., things to eat. Bhat et al. (2016) used only letter fluency, whereas two studies used both (Kim, 2021; Lee et al., 2021). In all these studies, the dual-task condition was motoric (e.g., walking, upper limb). In three studies, single word production involved picture naming (Bruehl et al., 2021; Laganaro et al., 2019) or phrase completion (Murray, 2000). In these studies, the competing task was an auditory-perceptual decision-making task.

Table 5 about here

Irrespective of type of task, statistically significant dual-task costs (i.e., worse performance in dual- as compared with single-task for within stroke participant comparisons) were found in five studies (Bhatt, 2016; Haggard et al., 2000; Laganaro et al., 2019; Lee et al., 2021; Murray, 2000); however, dual-task costs were not found in others (Cockburn et al., 2003; Feld & Plummer, 2021; Harley, 2006; Kim, 2021). In the three studies that used more complex dual-tasks (Tsang et al., 2019; Tsang & Pang, 2020; Yang et al., 2016), the results did not appear to be significant (95% CI pairwise estimates). In the case-series study by Bruehl et al. (2021), 11 of the 19 participants were less accurate in the dual-task condition.

Table 6 shows the methodological appraisal of single word studies. The overall quality ratings were low in all studies. Stroke and cognitive variables were the two categories where most studies received low ratings in comparison to the other two appraisal categories (study design, demographic variables).

Table 6 about here

Discourse production studies

Study characteristics and the individual results of the eight studies that reported discourse production data are shown in Tables 7 to 10. Table 11 shows these studies' critical appraisal ratings.

Table 7 about here

A total of 141 stroke participants were included in these eight studies (Table 7), with five studies including neurotypical controls. While age, education, and sex were reported in most studies, information about level of education was not reported in three studies (Plummer-D'Amato et al., 2008; Pohl et al., 2011a; Rogalski et al., 2010). Detailed cognitive profile information was included in most studies. Exceptions were the study by Pohl et al. (2011a) who provided cognitive information based on the MMSE. Murray et al. (1998) provided information about limb apraxia and estimated IQ. Only three studies reported information about motor speech disorders (Harmon et al., 2019; Plummer-D'Amato et al., 2008; Rogalski et al., 2010). In two studies, participants did not have severe dysarthria (Plummer-D'Amato et al., 2008; Rogalski et al., 2010); however, information about how these abilities were assessed was scarce, unlike in Harmon et al. (Harmon et al., 2019) who reported details about dysarthria and apraxia of speech assessment. Lastly, assessment of aphasia using standardised tests was evident in six of the eight studies. The participants in Rogalski et al. (2010) did not have aphasia according to a speech-language therapist's diagnosis. In Harmon et al. (2019) and Murray et al. (1998), all stroke participants had aphasia, whereas in Plummer et al. (2020), six of the 29 stroke participants had mild aphasia. The language skills of the remaining participants in Plummer et al. were within normal limits according to a bedside version of the

Western Aphasia Battery. Overall, across all eight discourse studies, 39 (of 141) stroke participants had aphasia.

Table 8 about here

Table 8 displays information about stroke. Type of stroke was reported in three of the eight studies (Plummer et al., 2020; Plummer-D'Amato et al., 2008; Rogalski et al., 2010). Similarly, five studies did not report information about site of lesion, whereas in Pohl et al. (2011a) this information was unclear because structural MRI data were provided for only 14 of the 19 participants (which included 8 left- and 6 right-hemisphere regions of interest). Kemper et al. (Kemper et al., 2006) included participants who had suffered right-, left-, or bilateral stroke. All participants in Murray et al. (1998) had a left-hemisphere stroke. Only two studies (Plummer et al., 2020; Pohl et al., 2011b) reported information on stroke severity, both using the Stroke Impact Scale (Duncan et al., 1999). Participants in most studies were in the chronic stage of post-stroke recovery, i.e., > 3 months post-stroke.

Table 9 about here

Information about task prioritisation (see Table 9) was included in all discourse production studies apart from Harmon et al. (2019). In five of the eight studies, participants were not to prioritise either activity. In Murray et al. (1998), participants were instructed to prioritise the discourse task (picture description), whereas Plummer et al. (2020) contrasted talking while prioritising walking in two settings (laboratory and real-world setting, which was a hospital lobby). Reliability information was provided in five of the eight studies, which was high across these studies.

Table 10 about here

The range of dual-tasks and dual-task effects are reported in Table 10. In six of the eight studies, the competing task in the dual-task condition involved motoric actions (walking or hand tapping). Receptive, tone discrimination tasks were used by Harmon et al. (2019) and Murray et al. (1998). Kemper et al. (Kemper et al., 2006), in addition to motoric tasks, also used receptive tasks (noise, speech) as part of their dual-task protocol. The language elicitation methods were predominantly narratives about personal opinions or general knowledge topics in six of the eight studies. Murray et al. used picture description, whereas Harmon et al. (Harmon et al., 2019) used story retelling, which was supported audio-visually and could be regarded as semi-spontaneous.

Dual-task effects were reported in most (7/8) studies but not necessarily in all spoken language measures. For example, both mild and moderate aphasia groups in Harmon et al. showed dual-task costs in number of correct information units (CIUs) and speech rate, but the ratio of CIUs per word showed dual-task costs only in the moderate aphasia group. Contrastingly, pauses per utterance showed dual-task costs in the mild group but not in the moderate group. In Kemper et al., speech rate, mean length of utterance, syntactic complexity, and propositional density were susceptible to dual-task costs across *all* dual-task conditions. In fact, only two language measures across the five dual-task conditions deviated from this pattern: the percentage of grammatical sentences was not affected when talking while ignoring speech, and type-token ratio in talking was not affected while walking. In Murray et al. (1998), dual-task costs in syntactic complexity were evident in the aphasia group in picture description and tone discrimination dual-task conditions; production of CIUs was also affected. Total word counts, percentage of word finding errors and unsuccessful sentences also showed dual-task effects in the aphasia group; verb morphology, however, was not affected. Mixed results were reported by Plummer-D'Amato et al. (2008): Number of utterances, words, and pauses showed dual-task costs whereas syntactic complexity and grammatical sentences did not. Plummer et al. (2020) reported dual-task costs regarding verb production. With regard to speech rate, mixed results (depending on the competing task) were reported in Pohl et al. (2011a), whereas in Pohl et al. (2011b), speech rate, mean length of utterance, and % of grammatical sentences were unaffected by the dual-task. Finally, Rogalski et al. (2010) did not find a dual-task effect in the two measures of discourse coherence they studied.

Table 11 about here

Table 11 shows the critical appraisal rating of the discourse production studies. Of note is that only Plummer et al. (Plummer et al., 2020) attained a high overall rating. All other studies were of low overall methodological quality, especially in the category of stroke-related variables with five studies getting a low rating (Harmon et al., 2019; Kemper et al., 2006; Pohl et al., 2011a; Pohl et al., 2011b; Rogalski et al., 2010).

Discussion

The main purpose of the present systematic review was to synthesise characteristics and findings of studies that compared language production performance in single- vs. dual-tasks among people affected by stroke of any severity. In this section, we discuss the key patterns that were evident in the data and their implications in accordance with the study's aims.

Our review identified 21 dual-task language production studies. Most focused on single word (n = 13) or discourse production (n = 8) and included 420 and 141 stroke participants, respectively (total of 561 stroke participants). In single word studies, word fluency was the main language task, being used in 10/13 studies. The remaining three studies, focused on single word naming and included people with aphasia (Bruehl et al., 2021; Laganaro et al., 2019; Murray, 2000). Also, in these three studies, the secondary perceptual task can be described as non-motoric, although it did require a button pressing response. Overall, dual-task costs were seen in these three studies. The other 10 single word studies involved competing motoric tasks; within these studies, dual-task costs were reported in five studies (Bhatt et al., 2016; Haggard et al., 2000; Harley et al., 2006; Tsang & Pang, 2020; Yang et al., 2016), whereas the other five studies did not (Cockburn et al., 2003; Feld & Plummer, 2021; Kim et al., 2021; Lee et al., 2021; Tsang et al., 2019). Two implications arise from these patterns: a) it is not clear if a perceptual secondary task creates dual-task effects in word production only in people with aphasia or the wider stroke population; and, b) it is also unclear which motoric tasks can interfere with word fluency in post-stroke cognitive impairment.

Our synthesis of tasks and findings shows a great variation among the semantic content of the word fluency tasks. Within the nine studies that used semantic fluency, different semantic categories were used during the dual-task condition, e.g., “things to eat” (Cockburn et al., 2003), or “hospital related words” (Kim, 2021). Some authors contrasted performance in “easier” versus “harder” semantic categories with seemingly statistically significant differences between them (Tsang et al., 2019). Additionally, within these nine studies there was variation in the dual-task, with most studies using walking, apart from two studies (Kim, 2021; Lee et al., 2021) who used an upper limb motoric task. In Lee et al., participants also carried out a phonemic word fluency task with data reported separately from semantic fluency. However, in the first of these studies, participants completed either a semantic or a phonemic fluency task, but the authors did not report data separately from each task. Word fluency studies of neurotypical older adults show that numerically, i.e., number of exemplars produced, scores in semantic fluency are larger than phonemic (Shao et al., 2014). Such findings point against combining scores at least numerically. One study (Bhatt, 2016) had their neurotypical participants complete a letter fluency task which only lasted 8 seconds unlike other studies which gave a longer time constraint (typically one minute). In appraising the presence or absence of dual-task language effects (negative or positive), consideration should also be given as to whether comparisons of performance from appropriate neurotypical controls were included. Only four verbal fluency studies reported data from controls (Bhatt, 2016; Haggard et al., 2000; Harley, 2006; Kim, 2021). Among them, two studies (Bhatt, 2016; Haggard et al., 2000) found dual-task cost differences between their stroke group and neurotypical controls; one study did not (Kim, 2021). However, one study (Harley, 2006) found dual-task effects in neurotypical controls but not in stroke participants.

Moving to discourse studies, the single- versus dual-task comparisons involved primarily walking or upper limb motoric tasks apart from two studies who used non-motoric, auditory discrimination tasks (Harmon et al., 2019; Murray et al., 1998). Neurotypical controls were included in five studies (Harmon et al., 2019; Kemper et al., 2006; Murray et al., 1998; Plummer et al., 2020; Pohl et al., 2011b). In terms of findings, dual-task costs in measures of semantic content, e.g., type-token ratio, correct CIUs, were reported in three of these studies (Kemper et al., 2006; Murray et al., 1998; Plummer et al., 2020). However, one study (Pohl et al., 2011b) did not find a difference in type-token ratio, whereas another (Harmon et al., 2019) found dual-task language cost differences only between controls and the moderate aphasia group, but not between controls and the mild aphasia group. Dual-task effects in terms of simplified syntactic complexity were reported in several studies (Kemper et al., 2006; Murray et al., 1998; Plummer et al., 2020; Pohl et al., 2011b). Considering the patterns of dual-task effects that were found across both types of studies (single word, discourse), it seems that dual-task costs are more likely to be present at the discourse as opposed to single word linguistic level.

Description of the participants' demographic backgrounds and cognitive-linguistic profiles was reported inconsistently. Of the 21 included studies in both categories, eight did not report important participant demographic information, i.e., age, education, and sex. Information about education was absent in eight studies, despite its well-documented influence on language production, word fluency (Jebahi et al., 2020) and discourse (Le Dorze & Bédard, 1998). In terms of cognitive-linguistic variables, most single word and discourse studies used at least one other cognitive test beyond language to describe cognitive abilities. All discourse studies provided some information about cognition beyond language. However, across both types of studies, there was range of cognitive tests with the MMSE and MoCA being the most popular. These tests were used either for inclusion/exclusion purposes without providing statistical summaries (Kim, 2021) or for characterising participants' cognitive abilities via descriptive statistics (Rogalski et al., 2010). Beyond these two tests, short-term memory, vocabulary, and digit-symbol processing speed tests were used more than once to describe cognitive profiles of the samples. Such inconsistency makes comparison across studies difficult. Additionally, scarcity of information about key cognitive abilities presents problems in determining with more precision the causes of dual-task decrements, for example, executive function, attention, and task complexity.

There was limited information about post-stroke communication disorders (i.e., dysarthria, apraxia of speech, aphasia). As speech is the vehicle of language and given the prevalence of motor speech disorders (dysarthria and/or apraxia of speech) post-stroke (Ali et al., 2015), this information is crucial to include. However, Harmon et al. (2019) presented exemplary detail for both motor speech disorders. Relatedly, information about the language abilities of participants (via an aphasia test) was included in less than half of the studies, i.e., n = 10, with five of these focusing specifically on aphasia. The total number of people with aphasia was 74 (of 561 in total). Without comprehensive description of motor speech and language abilities, comparisons of findings across studies become difficult, and the extent to which dual-task language effects stem from language abilities as opposed to a combination of language, speech and divided attention abilities remains unclear.

Stroke-related information was not reported systematically. Only eight of the 21 studies included some stroke-related information such as gross lesion site (left-right); the other 13 did not.

Most studies included individuals with both left- or right-hemisphere stroke. Stroke type (ischaemic, haemorrhagic) was reported scantly. Thus, the extent to which gross lesion site and stroke type may influence dual-task language performance is largely unknown. There was only one study (Murray, 2000) that compared dual-task language performance between individuals with left- or right-hemisphere stroke. Two aphasia-specific studies exclusively included participants with left-hemisphere stroke (Bruehl et al., 2021; Laganaro et al., 2019). While language is considered primarily a left-hemisphere ability, different language tasks engage both hemispheres (Biesbroek et al., 2016; Bradshaw et al., 2017). Stroke severity was also scarcely reported. Notably only two studies (Feld & Plummer, 2021; Plummer et al., 2020) included some information about stroke severity. In summary, the overall cursory description of the participant samples thus far makes it challenging to determine if certain individuals after stroke are more or less vulnerable to dual-task effects; additionally, without the inclusion of well-matched neurotypical controls, whether dual-task effects are related to stroke or typical ageing cannot be determined.

Overall, little is known about the effect of task prioritisation and dual-task language effects in the stroke literature. Of the 15 studies (across single word and discourse studies) that included information about task prioritisation, nine instructed participants explicitly not to prioritise one task over another or perform both tasks equally well; the remaining six studies included information about task prioritisation but differed in whether participants were instructed to prioritise the language or competing task. Across studies, such variation in or lack of information pertaining to prioritisation further confounds understanding when post-stroke language abilities may be vulnerable during dual-task conditions. Furthermore, specification of task prioritisation instructions is critical to examining if executive attention allocation issues are contributing to dual-task effects (Kahneman, 1973); that is, without knowing the prioritisation instructions, it cannot be determined whether performance of one or both tasks might be expected to demonstrate dual-task effects and whether the participants with stroke were able to prioritize as instructed.

Our certainty (or confidence) assessment was based on the methodological appraisal of the studies and information about reliability/fidelity. All 13 single word studies attained a low overall rating, with only four reporting information about reliability/fidelity. Inter- or intra-rater agreement was reported in two studies (Laganaro et al., 2019; Murray, 2000). Although test-retest reliability was reported in other studies (Tsang et al., 2019; Yang et al., 2016), these studies did not report as to whether reliability of raw data coding was carried out. Similarly, seven of the eight discourse studies were of low quality, the exception being Plummer et al. (Plummer et al., 2020) which received a high overall rating. However, no reliability/fidelity analyses were carried out in this study. In discourse studies in which reliability/fidelity information was reported, this was high (> .80) for both intra- and inter-rater aspects. Considering the number of studies in both single word and discourse categories (n = 10) that included appropriate control groups and the inconsistent findings with regard to presence/absence of dual-task effects across studies, the certainty of the findings in this segment of the stroke literature may be described as weak. We should note that issues of reliability in dual-task studies are evident in other populations (Pike et al., 2022).

Moving forward, the following recommendations for future research may be made. The methodological rigour should be improved by including appropriate neurotypical controls of comparable age, level of education, and sex to stroke participants in between-subject experimental designs; larger sample sizes to maximise both internal and external validity of the findings are also needed. It is critical that future studies adopt a more consistent approach to describing important participant characteristics, e.g., education, to enable a more meaningful understanding of the evidence-base through cross-study comparisons (Stark et al., 2022; Wallace et al., 2022). Likewise, there needs to be a more consistent and detailed approach to documenting and describing motor speech, language, and broader cognitive skills, given the impact these skills may have on spoken language task performance, particularly within the dual-task paradigm. Additionally, individual data of all participant subgroups (stroke, controls) should be accessible in supplementary materials because not all participants show dual-task effects.

In most studies, the type of task combinations to discern language-specific dual-task effects was competing spoken language and motoric tasks. Nonetheless, in some studies the dual-task involved upper limb movement while talking, whereas in others lower limb movement while talking. Therefore, the possible influence of these competing task variables on language within the dual-task paradigm remains unknown. For example, only in Kemper et al. (Kemper et al., 2006) did participants engaged both in motoric as well as receptive listening tasks. Their findings showed that dual-task effects were evident in most (though not all) language production measures when the competing task was motoric or receptive.

Delineating the role of task within the current dual-task findings is compounded by our limited understanding of how task prioritisation instructions may affect language behaviour; that is, in most studies that reported task prioritisation information, participants were not told to prioritise one task over the other. Understanding the effect of task prioritisation is important as it could have treatment implications, especially provision of advice in everyday communicative environments. In summary, the influence of task prioritisation instructions on dual-task language effects is largely unknown. Consequently, systematic examination of varying task prioritization instructions to determine attention allocation abilities is needed.

Another major gap is that brain-behaviour relationships in language dual-tasking after stroke are also largely unknown. We identified only one study (Pohl et al., 2011a) that included structural MRI lesion mapping for some of its participants, although its primary purpose was to ascertain extent of stroke lesion rather than investigate brain-behaviour relations. More of a concern is the limited information about gross lesion location (right-left) of stroke in the included studies.

However, we should also note that apart from one study (Murray, 2000), none of the included studies had the investigation of brain-behaviour relationships as their primary goal. Given the functional lateralisation complexities of attention, language, and executive functions as well as motoric abilities (Baldassarre et al., 2016; Biesbroek et al., 2016; Lehman Blake, 2018) this is a major void in the current knowledge of language-specific dual-task effects.

We found no study that focused explicitly on TIA. Although historically the symptoms of TIAs were thought to disappear within 24 hours (Caplan, 2006), recent evidence suggests that mild cognitive impairment is present for longer in TIA patients and its profile is similar to post-stroke cognitive impairment (van Rooij et al., 2017). Further, there is a clinical entity involving people after stroke who may or may not have demonstrated aphasia in the acute or subacute stage and/or subsequently recovered from aphasia as defined by within normal limits performance on diagnostic aphasia test (Salis & DeDe, 2022; Vallar et al., 1988). Despite its limitations, Kemper and colleagues (Kemper et al., 2006) conducted the only study that focused explicitly on the diagnostic potential of language-specific dual-tasks (elicited from discourse production) in discerning post-stroke latent language problems to which standardised aphasia tests are insensitive. Feld and Plummer (Feld & Plummer, 2021) studied individuals whose language performance was overall within normal limits according to an aphasia test. However, their findings did not show diminished word fluency performance under a dual-task condition, at least at the group level. In both these studies, participants with left and right hemisphere strokes were included. Accordingly, there is scope for further research into the diagnostic potential of the dual-task paradigm in identifying latent language deficits.

The main limitation in the present review is its focus on peer reviewed studies and not on “grey” literature. Therefore, we may not have included relevant literature and consequently introduced reporting or publication bias (Paez, 2017; Tricco et al., 2018). Two studies (Murray, 2000;

Murray et al., 1998) that were included in this review were conducted by one of the authors. While this author was not included in the data extraction or appraisal of these studies, the possibility of unbiased appraisal cannot be excluded. Also, we were only able to review studies published in English. Finally, we did not contact authors to request information that was not included in their studies.

To conclude, in this systematic review of language-specific dual-task effects in stroke, we found that more studies and with greater number of participants focused on single word production than studies that focused on discourse. In terms of dual-task effects, language-specific dual-task costs were identified in single word studies, especially those that focused on people with aphasia as well as half of the non-aphasia studies. Unlike the single word studies, however, nearly all studies of discourse showed dual-task decrements on at least some variables. These findings should be predicated by the following shortcoming. We raised concerns about the absence of appropriate data from neurotypical controls, limited understanding of the influence of left versus right lesion site, as well as the demands of task prioritisation instructions on dual-task language performance. Additionally, more detailed and consistent descriptions of participants' demographic variables as well as motor speech, language and broader cognitive skills should be reported. These limitations point clearly to avenues for further research to improve our understanding of language-specific dual-task effects after stroke.

Data Availability Statement: The original data are available from the corresponding author.

References

Ali, M., Lyden, P., & Brady, M. (2015). Aphasia and Dysarthria in Acute Stroke: Recovery and Functional Outcome. *International Journal of Stroke*, *10*, 400–406. https://doi.org/10.1111/ijs.12067

Al-Yahya, E., Dawes, H., Smith, L., Dennis, A., Howells, K., & Cockburn, J. (2011). Cognitive motor interference while walking: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *35*(3), 715–728. https://doi.org/10.1016/j.neubiorev.2010.08.008

Babbitt, E. M., Heinemann, A. W., Semik, P., & Cherney, L. R. (2011). Psychometric properties of the Communication Confidence Rating Scale for Aphasia (CCRSA): Phase 2. *Aphasiology*, *25*, 727–735. https://doi.org/10.1080/02687038.2010.537347

Baldassarre, A., Ramsey, L., Rengachary, J., Zinn, K., Siegel, J. S., Metcalf, N. V., Strube, M. J., Snyder, A. Z., Corbetta, M., & Shulman, G. L. (2016). Dissociated functional connectivity profiles for motor and attention deficits in acute right-hemisphere stroke. *Brain*, *139*(7), 2024–2038. https://doi.org/10.1093/brain/aww107

Bhatt, T. (2016). Examining interference of different cognitive tasks on voluntary balance control in aging and stroke. *Experimental Brain Res Research*, 10.

Biesbroek, J. M., van Zandvoort, M. J. E., Kappelle, L. J., Velthuis, B. K., Biessels, G. J., & Postma, A. (2016). Shared and distinct anatomical correlates of semantic and phonemic fluency revealed by lesion-symptom mapping in patients with ischemic stroke. *Brain Structure and Function*, *221*(4), 2123–2134. https://doi.org/10.1007/s00429-015-1033-8

Blake, H. (2002). An evaluation of screening measures for cognitive impairment after stroke. *Age and Ageing*, *31*(6), 451–456. https://doi.org/10.1093/ageing/31.6.451

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Lijmer, J. G., Moher, D., & Rennie, D. (2003). *Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative*. *326*, 4.

Bradshaw, A. R., Thompson, P. A., Wilson, A. C., Bishop, D. V. M., & Woodhead, Z. V. J. (2017). Measuring language lateralisation with different language tasks: A systematic review. *PeerJ*, *5*, e3929. https://doi.org/10.7717/peerj.3929

Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews*, *2016*(6). https://doi.org/10.1002/14651858.CD000425.pub4

Brott, T., Adams, H. P., Olinger, C. P., Marler, J. R., Barsan, W. G., Biller, J., Spilker, J., Holleran, R., Eberle, R., & Hertzberg, V. (1989). Measurements of acute cerebral infarction: A clinical examination scale. *Stroke*, *20*, 864–870. https://doi.org/10.1161/01.STR.20.7.864

Bruehl, S., Willmes, K., & Binkofski, F. (2021). Interfered-Naming Therapy for Aphasia (INTA): Behavioural and computational effects of a novel linguistic-executive approach. *Aphasiology*, 1–22. https://doi.org/10.1080/02687038.2021.1995841

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., & Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ*, l6890. https://doi.org/10.1136/bmj.l6890

Caplan, L. R. (2006). Transient ischemic attack: Definition and natural history. *Current Atherosclerosis Reports*, *8*, 276–280.

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The Ecological Validity of Neuropsychological Tests: A Review of the Literature on Everyday Cognitive Skills. *Neuropsychology Review*, 17.

Cockburn, J., Haggard, P., Cock, J., & Fordham, C. (2003). Changing patterns of cognitive-motor interference (CMI) over time during recovery from stroke. *Clinical Rehabilitation*, *17*, 167–173.

Cockburn, J., Haligan, P. W., & Wilson, B. A. (1987). *Behavioral Inattention Test*. Pearson.

Cristofori, I., & Levine, H. S. (2015). Traumatic brain injury and cognition. In *Handbook of Clinical Neurology* (Vol. 128, pp. 579–611). Elsevier.

Deblock-Bellamy, A., Lamontagne, A., & Blanchette, A. K. (2020). Cognitive-Locomotor Dual-Task Interference in Stroke Survivors and the Influence of the Tasks: A Systematic Review. *Frontiers in Neurology*, *11*, 882. https://doi.org/10.3389/fneur.2020.00882

DeDe, G., & Salis, C. (2020). Temporal and Episodic Analyses of the Story of Cinderella in Latent Aphasia. *American Journal of Speech-Language Pathology*, *29*, 449–462.

Dignen, B. (2009). *Systematic Reviews: CRD's guidance for undertaking reviews in health care*. York Associates International.

Duncan, P. W., Wallace, D., Lai, S. M., Johnson, D., Embretson, S., & Laster, L. J. (1999). The Stroke Impact Scale Version 2.0: Evaluation of Reliability, Validity, and Sensitivity to Change. *Stroke*, *30*(10), 2131–2140. https://doi.org/10.1161/01.STR.30.10.2131

Feld, J. A., & Plummer, P. (2021). Patterns of cognitive-motor dual-task interference post stroke: An observational inpatient study at hospital discharge. *European Journal of Physical and Rehabilitation Medicine*, *57*(3), 327–336.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Ghai, S., Ghai, I., & Effenberg, A. O. (2017). Effects of dual tasks and dual-task training on postural stability: A systematic review and meta-analysis. *Clinical Interventions in Aging*, *Volume 12*, 557–577. https://doi.org/10.2147/CIA.S125201

Gowland, C., Stratford, P., Ward, M., Moreland, J., Torresin, W., Van Hullenaar, S., Sanford, J., Barreca, S., Vanspall, B., & Plews, N. (1993). Measuring physical impairment and disability with the Chedoke-McMaster Stroke Assessment. *Stroke*, *24*, 58–63. https://doi.org/10.1161/01.STR.24.1.58

Haggard, P., Cockburn, J., Cock, J., Fordham, C., & Wade, D. (2000). Interference between gait and cognitive tasks in a rehabilitating neurological population. *Journal of Neurology, Neurosurgery, and Psychiatry*, *69*, 479–486.

Harley, C. (2006). Disruption of sitting balance after stroke: Influence of spoken output. *Journal of Neurology, Neurosurgery & Psychiatry*, *77*(5), 674–676. https://doi.org/10.1136/jnnp.2005.074138

Harmon, T. G., Jacks, A., Haley, K. L., & Bailliard, A. (2019). Dual-Task Effects on Story Retell for Participants With Moderate, Mild, or No Aphasia: Quantitative and Qualitative Findings. *Journal of Speech, Language, and Hearing Research*, *62*, 1890–1905.

He, Y., Yang, L., Zhou, J., Yao, L., & Pang, M. Y. C. (2018). Dual-task training effects on motor and cognitive functional abilities in individuals with stroke: A systematic review. *Clinical Rehabilitation*, *32*(7), 865–877. https://doi.org/10.1177/0269215518758482

Helm-Estabrooks, N. (1992). *Aphasia Diagnostic Profiles*. Pro-Ed.

Hillis, A. E., & Tippett, D. C. (2014). Stroke Recovery: Surprising Influences and Residual Consequences. *Advances in Medicine*, *2014*, 1–10. https://doi.org/10.1155/2014/378263

Hofheinz, M., Mibs, M., & Elsner, B. (2016). Dual task training for improving balance and gait in people with stroke. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD012403

Hoppe, C., Muller, U., Weheid, K., Thone, A., & von Cramon. (2000). Digit Ordering Test: Clinical, psychometric, and experimental evaluation of a verbal working memory test. *Clinical Neuropsychologist*, *14*, 38–55.

Hula, W., & McNeil, M. (2008). Models of Attention and Dual-Task Performance as Explanatory Constructs in Aphasia. *Seminars in Speech and Language*, *29*(03), 169–187. https://doi.org/10.1055/s-0028-1082882

Jaywant, A., Toglia, J., Gunning, F. M., & O'Dell, M. W. (2019). The diagnostic accuracy of the Montreal Cognitive Assessment in inpatient stroke rehabilitation. *Neuropsychological Rehabilitation*, *29*(8), 1163–1176. https://doi.org/10.1080/09602011.2017.1372297

Jebahi, F., Abou Jaoude, R., & Ellis, C. (2020). Semantic verbal fluency task: The effects of age, educational level, and sex in Lebanese-speaking adults. *Applied Neuropsychology: Adult*, 1–5. https://doi.org/10.1080/23279095.2020.1821031

Jokinen, H., Melkas, S., Ylikoski, R., Pohjasvaara, T., Kaste, M., Erkinjuntti, T., & Hietanen, M. (2015). Post-stroke cognitive impairment is common even after successful clinical recovery. *European Journal of Neurology*, *22*(9), 1288–1294. https://doi.org/10.1111/ene.12743

Kahneman, D. (1973). *Attention and Effort*. Prentice Hall.

Kemper, S., McDowd, J., Pohl, P., Herman, R., & Jackson, S. (2006). Revealing Language Deficits Following Stroke: The Cost of Doing Two Things at Once. *Aging, Neuropsychology, and Cognition*, *13*(1), 115–139. https://doi.org/10.1080/13825580500501496

Kertesz, A. (2007). *Western Aphasia Battery-Revised (WAB-R)*. Pearson.

Kim, H. (2021). Dual Task Effects on Speed and Accuracy During Cognitive and Upper Limb Motor Tasks in Adults With Stroke Hemiparesis. *Frontiers in Human Neuroscience*, *15*, 671541.

Kontou, E., Walker, M., Thomas, S., Watkins, C., Griffiths, H., Golding-Day, M., Richardson, C., & Sprigg, N. (2020). Optimising Psychoeducation for Transient Ischaemic Attack and Minor Stroke Management (OPTIMISM): Protocol for a feasibility randomised controlled trial. *AMRC Open Research*, *2*, 24. https://doi.org/10.12688/amrcopenres.12911.1

Laganaro, M., Bonnans, C., & Fargier, R. (2019). Word form encoding is under attentional demand: Evidence from dual-task interference in aphasia. *Cognitive Neuropsychology*, *36*, 18–30.

Lazar, R. M., & Boehme, A. K. (2017). Aphasia As a Predictor of Stroke Outcome. *Current Neurology and Neuroscience Reports*, *17*(11), 83. https://doi.org/10.1007/s11910-017-0797-z

Le Dorze, G., & Bédard, C. (1998). Effects of age and education on the lexico-semantic content of connected speech in adults. *Journal of Communication Disorders*, *31*(1), 53–71. https://doi.org/10.1016/S0021-9924(97)00051-8

Lee, K. J., Park, G., & Shin, J.-H. (2021). Differences in Dual Task Performance After Robotic Upper Extremity Rehabilitation in Hemiplegic Stroke Patients. *Frontiers in Neurology*, *12*, 771185. https://doi.org/10.3389/fneur.2021.771185

Lee, Y. J., & Jung, M. Y. (2016). A systematic review of the dual-task training for stroke with hemiplegia. *Therapeutic Science for Rehabilitation*, *5*, 23–32.

Lehman Blake, M. (2018). *The Right Hemisphere and Disorders of Cognition and Communication Theory and Clinical Practice*. Plural.

Lo, J. W., Crawford, J. D., Desmond, D. W., Godefroy, O., Jokinen, H., Mahinrad, S., Bae, H.-J., Lim, J.-S., Köhler, S., Douven, E., Staals, J., Chen, C., Xu, X., Chong, E. J., Akinyemi, R. O., Kalaria, R. N., Ogunniyi, A., Barbay, M., Roussel, M., … for the Stroke and Cognition (STROKOG) Collaboration. (2019). Profile of and risk factors for poststroke cognitive impairment in diverse ethnoregional groups. *Neurology*, *93*(24), e2257–e2271. https://doi.org/10.1212/WNL.0000000000008612

Loetscher, T., Potter, K.-J., Wong, D., & das Nair, R. (2019). Cognitive rehabilitation for attention deficits following stroke. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD002842.pub3

Martin, N., Kohen, F., Kalinyak-Fliszar, M., Soveri, A., & Laine, M. (2012). Effects of working memory load on processing of sounds and meanings of words in aphasia. *Aphasiology*, *26*, 462–493. https://doi.org/10.1080/02687038.2011.619516

Murray, L. L. (2000). The Effects of Varying Attentional Demands on the Word Retrieval Skills of Adults with Aphasia, Right Hemisphere Brain Damage, or No Brain Damage. *Brain and Language*, *72*, 40–72.

Murray, L. L. (2012). Attention and Other Cognitive Deficits in Aphasia: Presence and Relation to Language and Communication Measures. *American Journal of Speech-Language Pathology*, *21*(2), 51–64. https://doi.org/10.1044/1058-0360(2012/11-0067)

Murray, L. L., & Coppens, P. (2021). Formal and informal assessment of aphasia. In *Aphasia and related neurogenic communication disorders* (3rd.). Jones & Bartlett.

Murray, L. L., Holland, A. L., & Beeson, P. M. (1998). *Spoken Language of Individuals With Mild Fluent Aphasia Under Focused and Divided-Attention Conditions*. *41*, 15.

Murray, L., Salis, C., Martin, N. ., & Dralle, J. (2018). The use of standardised short-term and working memory tests in aphasia research: A systematic review. *Neuropsychological Rehabilitation*, *28*, 309–351.

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). *The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment*. *53*(4), 5.

Nelson, H. E. (1982). *The National Adult Reading Test (NART): Test manual*. NFER-Nelson.

Nicolas, K., Levi, C., Evans, T.-J., Michie, P. T., Magin, P., Quain, D., Bivard, A., & Karayanidis, F. (2020). Cognition in the First Year After a Minor Stroke, Transient Ischemic Attack, or Mimic Event and the Role of Vascular Risk Factors. *Frontiers in Neurology*, *11*, 216. https://doi.org/10.3389/fneur.2020.00216

Oberauer, K. (2019). *Working Memory and Attention – A Conceptual Analysis and Review*. 23.

Paez, A. (2017). Gray literature: An important resource in systematic reviews. *Journal of Evidence-Based Medicine*, *10*(3), 233–240. https://doi.org/10.1111/jebm.12266

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71

Pashler, H. (1994). Dual-task interference in simple tasks. *Psychological Bulletin*, *116*, 220–244.

Pfeiffer, E. (1975). A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *Journal of the American Geriatrics Society*, *23*, 433–441.

Pike, A., McGuckian, T. B., Steenbergen, B., Cole, M. H., & Wilson, P. H. (2022). How Reliable and Valid are Dual-Task Cost Metrics? A Meta-Analysis of Locomotor-Cognitive Dual-Task Paradigms.

Archives of Physical Medicine and Rehabilitation, S000399932200572X. https://doi.org/10.1016/j.apmr.2022.07.014

Pimental, P., & Kingsbury, N. (1989). *Mini Inventory of Right Brain Injury*. Pro-Ed.

Plummer, P., Altmann, L. J. P., Feld, J., Zukowski, L., Najafi, B., & Giuliani, C. (2020). Attentional prioritization in dual-task walking_ Effects of stroke, environment, and instructed focus. *Gait & Posture*, *79*, 3–9.

Plummer, P., & Eskes, G. (2015). Measuring treatment effects on dual-task performance: A framework for research and clinical practice. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00225

Plummer, P., Eskes, G., Wallace, S., Giuffrida, C., Fraas, M., Campbell, G., Clifton, K.-L., & Skidmore, E. R. (2013). Cognitive-Motor Interference During Functional Mobility After Stroke: State of the Science and Implications for Future Research. *Archives of Physical Medicine and Rehabilitation*, *94*(12), 2565-2574.e6. https://doi.org/10.1016/j.apmr.2013.08.002

Plummer, P., & Iyigün, G. (2018). Effects of Physical Exercise Interventions on Dual-Task Gait Speed Following Stroke: A Systematic Review and Meta-Analysis. *Archives of Physical Medicine and Rehabilitation, 99*, 2548–2560.

Plummer-D'Amato, P., Altmann, L. J. P., Saracino, D., Fox, E., Behrman, A. L., & Marsiske, M. (2008). *Interactions between cognitive tasks and gait after stroke: A dual task study*. 6.

Pohl, P. S., Kemper, S., Siengsukon, C. F., Boyd, L., Vidoni, E. D., & Herman, R. E. (2011a). Dual-Task Demands of Hand Movements for Adults with Stroke: A Pilot Study. *Topics in Stroke Rehabilitation*, *18*, 238–247. https://doi.org/10.1310/tsr1803-238

Pohl, P. S., Kemper, S., Siengsukon, C. F., Boyd, L., Vidoni, E., & Herman, E. (2011b). Older Adults With and Without Stroke Reduce Cadence to Meet the Demands of Talking. *Journal of Geriatric Physical Therapy*, *34*, 6.

Poulin, V., Korner-Bitensky, N., Dawson, D. R., & Bherer, L. (2012). Efficacy of Executive Function Interventions After Stroke: A Systematic Review. *Topics in Stroke Rehabilitation*, *19*(2), 158–171. https://doi.org/10.1310/tsr1902-158

Randolph, C. (2012). *Repeatable Battery for the Assessment of Neuropsychological Status Update*. Pearson.

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1996). The structure of normal human attention: The Test of Everyday Attention. *Journal of the International Neuropsychological Society*, *2*(6), 525–534. https://doi.org/10.1017/S1355617700001697

Rogalski, Y., Altmann, L. J. P., Plummer-D'Amato, P., Behrman, A. L., & Marsiske, M. (2010). Discourse coherence and cognition after stroke: A dual task study. *Journal of Communication Disorders*, *43*(3), 212–224. https://doi.org/10.1016/j.jcomdis.2010.02.001

Salis, C., & DeDe, G. (2022). Sentence Production in a Discourse Context in Latent Aphasia: A Real-Time Study. *American Journal of Speech-Language Pathology*, *31*, 1284–1296.

Salis, C., Murray, L., & Vonk, J. M. J. (n.d.). *Systematic review of subjective memory measures to inform assessing memory limitations after stroke and stroke-related aphasia*. 20.

Salis, C., Murray, L., & Vonk, J. M. J. (2021). Systematic review of subjective memory measures to inform assessing memory limitations after stroke and stroke-related aphasia. *Disability and Rehabilitation*, *43*(11), 1488–1506. https://doi.org/10.1080/09638288.2019.1668485

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00772

Shewan, C. M., & Kertesz, A. (1980). *Reliability and validity characteristics of the Western Aphasia Battery (WAB)*. *45*, 308–324.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127–190.

Sohlberg, M. M., & Mateer, C. (2001). *Cognitive rehabilitation: An integrated neuropsychological approach*. Guilford.

Spaccavento, S., Marinelli, C. V., Nardulli, R., Macchitella, L., Bivona, U., Piccardi, L., Zoccolotti, P., & Angelelli, P. (2019). Attention Deficits in Stroke Patients: The Role of Lesion Characteristics, Time from Stroke, and Concomitant Neuropsychological Deficits. *Behavioural Neurology*, *2019*, 1–12. https://doi.org/10.1155/2019/7835710

Stark, B. C., Bryant, L., Themistocleous, C., den Ouden, D.-B., & Roberts, A. C. (2022). Best practice guidelines for reporting spoken discourse in aphasia and neurogenic communication disorders. *Aphasiology*, 1–24. https://doi.org/10.1080/02687038.2022.2039372

Stolwyk, R. J., O'Neill, M. H., McKay, A. J. D., & Wong, D. K. (2014). Are Cognitive Screening Tools Sensitive and Specific Enough for Use After Stroke?: A Systematic Literature Review. *Stroke*, *45*(10), 3129–3134. https://doi.org/10.1161/STROKEAHA.114.004232

Strobach, T., Wendt, M., & Janczyk, M. (2018). Editorial: Multitasking: Executive Functioning in Dual-Task and Task Switching Situations. *Frontiers in Psychology*, *9*, 108. https://doi.org/10.3389/fpsyg.2018.00108

Stroop, J. R. (1935). Studies of interference in serial verbal reaction. *Journal of Experimental Psychology*, *18*, 643–662.

Tang, E. Y., Amiesimaka, O., Harrison, S. L., Green, E., Price, C., Robinson, L., Siervo, M., & Stephan, B. C. (2018). *Longitudinal Effect of Stroke on Cognition: A Systematic Review*. 23.

Tatemichi, T. K., Desmond, D. W., Stern, Y., Paik, M., Sano, M., & Bagiella, E. (1994). Cognitive impairment after stroke: Frequency, patterns, and relationship to functional abilities. *Journal of Neurology, Neurosurgery & Psychiatry*, *57*(2), 202–207. https://doi.org/10.1136/jnnp.57.2.202

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA Extension for Scoping Reviews

(PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850

Tsang, C. S. L., Chong, D. Y. K., & Pang, M. Y. C. (2019). Cognitive-motor interference in walking after stroke: Test–retest reliability and validity of dual-task walking assessments. *Clinical Rehabilitation*, *33*(6), 1066–1078. https://doi.org/10.1177/0269215519828146

Tsang, C. S. L., & Pang, M. Y. C. (2020). Association of subsequent falls with evidence of dual-task interference while walking in community-dwelling individuals after stroke. *Clinical Rehabilitation*, *34*(7), 971–980. https://doi.org/10.1177/0269215520923700

Tsang, C. S.-L., Wang, S., Miller, T., & Pang, M. Y.-C. (2022). Degree and pattern of dual-task interference during walking vary with component tasks in people after stroke: A systematic review. *Journal of Physiotherapy*, *68*(1), 26–36. https://doi.org/10.1016/j.jphys.2021.12.009

Turner, G. M., McMullan, C., Atkins, L., Foy, R., Mant, J., & Calvert, M. (2019). TIA and minor stroke: A qualitative study of long-term impact and experiences of follow-up care. *BMC Family Practice*, *20*(1), 176. https://doi.org/10.1186/s12875-019-1057-x

Vallar, G., Papagno, C., & Cappa, S. F. (1988). Latent dysphasia after left hemisphere lesions: A lexical-semantic and verbal memory deficit. *Aphasiology*, *2*, 463–478.

van Rooij, F. G., Plaizier, N. O., Vermeer, S. E., Góraj, B. M., Koudstaal, P. J., Richard, E., de Leeuw, F.-E., Kessels, R. P. C., & van Dijk, E. J. (2017). Subjective Cognitive Impairment, Depressive Symptoms, and Fatigue after a TIA or Transient Neurological Attack: A Prospective Study. *Behavioural Neurology*, *2017*, 1–7. https://doi.org/10.1155/2017/5181024

Wallace, S. J., Isaacs, M., Ali, M., & Brady, M. C. (2022). Establishing reporting standards for participant characteristics in post-stroke aphasia research: An international e-Delphi exercise and consensus meeting. *Clinical Rehabilitation*.

Wang, X. -Q., Pi, Y. -L., Chen, B. -L, Chen, P. -J., Liu, Y., Wang, R., Li, X., & Waddington, G. (2015). Cognitive motor interference for gait and balance in stroke: A systematic review and meta-analysis. *European Journal of Neurology*, *22*(3), 555. https://doi.org/10.1111/ene.12616

Wechsler, D. (1958). *The Measurement and Appraisal of Adult Intelligence* (4th ed.). William & Wilkins.

Wechsler, D. (2009). *Wechsler Memory Scale*. Pearson.

Wecshler, D. (1999). *Manual for the Wechsler Abbreviated Scale of Intelligence*. Psychological Corporation.

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal Fluency: Language or Executive Function Measure? *Applied Neuropsychology: Adult*, *23*(1), 29–34. https://doi.org/10.1080/23279095.2015.1004574

Won, K. A., Lim, S. J., Park, H. Y., & Park, J. H. (2020). Effects of the Dual-Task Training on Stroke Patients: A Systematic Review and Meta-analysis. *Therapeutic Science for Rehabilitation*, *9*, 7–25.

Wong, M. T. P., Ho, T. P., Ho, M. Y., Yu, C. S., Wong, Y. H., & Lee, S. Y. (2002). Development and inter-rater reliability of a standardized verbal instruction manual for the Chinese Geriatric Depression Scale - short form. *International Journal of Geriatric Psychiatry*, *17*(5), 459–463. https://doi.org/10.1002/gps.633

Woollacott, M., & Shumway-Cook, A. (2002). Attention and the control of posture and gait: A review of an emerging area of research. *Gait and Posture*, *16*, 1–14.

Yang, L., He, C., & Pang, M. Y. C. (2016). Reliability and Validity of Dual-Task Mobility Assessments in People with Chronic Stroke. *PLOS ONE*, 22.

Zhou, Q., Yang, H., Zhou, Q., & Pan, H. (2021). Effects of cognitive motor dual-task training on stroke patients: A RCT-based meta-analysis. *Journal of Clinical Neuroscience*, *92*, 175–182. https://doi.org/10.1016/j.jocn.2021.08.009

Figure 1. Identification-inclusion process

Supplementary Materials

S1 PRISMA Abstracts checklist
S2 PRISMA Main checklist
S3 SWiM checklist
S4 Details of database searches
S5 Study appraisal tool

Table 1 Search terms used in electronic searches

population	task	language
aphasia, cerebrovascular accident, cerebrovascular-accident, CVA, dysphasia, mild stroke, mini stroke, mini-stroke, minor stroke, stroke, TIA, transient ischaemic attack, transient ischemic attack	cognitive control, concurrent task, distract*, divided attention, dual task*, dual-task*, interference, multitask*, multi-task*, selective attention	communication, conversation, discourse, language, linguistic, naming, sentence*, speak*, speech, spoken, word*

Table 2 Single word studies – Participant information

Studies	Participants	Demographics	Cognitive-linguistic variables	Motor speech disorder information	Information about aphasia or language ability
Bhatt et al. (2016)	10 stroke; 10 neurotypical older adults (OA); 10 neurotypical younger adults (YA)	*Stroke: mean age* = 57.2 (SD = 7.17); *Sex* = 6 M, 4 F; *OA: mean age* = 61 (SD = 5.53); *Sex* = 6 M, 4 F; *YA: mean age* = 25.54 (SD = 4.13); *Education:* not reported *Sex* = 4 M, 6 F	Not reported	Not reported	No aphasia according to physician's confirmation
Bruehl et al. (2021)	19 stroke	*Stroke: mean age* = 47, range = 31 – 74; *Mean education:* primary or secondary = 11.8 years (range 9 – 13); also 8 participants either attended or graduated from university *Sex* = 14 M, 7 F	Spoken naming performance	Apraxia of speech in 12 of the 20 participants	All participants presented with aphasia
Cockburn et al. (2003)	10 stroke	*Mean age* = 58, range = 41 – 71; *Education:* not reported *Sex* = 5 M, 5 F	NART; WAIS immediate and delayed story recall, digit span (forward, backward); letter fluency	Not reported	Sufficient language to perform the cognitive tasks
Feld & Plummer (2021)	47 stroke	*Mean age* = 59.5 (SD = 11.7) *Mean education* = 14 (SD = 2.8)	MoCA; Bedside WAB	Not reported	Mean Aphasia Quotient on Bedside WAB = 97.4 (i.e., within normal limits)

		Sex = 28 M, 19 F			
Haggard et al. (2000))	33 stroke (of a total of 50); 10 neurotypical controls	*Stroke: Mean age* = 50.18 (SD = 16.47), range = 18 – 84; *Education:* not reported *Sex* = 28 M, 22 F	Word fluency; mental arithmetic; paired associate monitoring; visuospatial decision	Not reported	Adequate language comprehension and production to follow instructions
Harley et al. (2006)	36 stroke; 21 neurotypical controls	*Stroke: mean age* = 61.6 years (SD = 15.9); *OA: mean age* = 71.0 years (SD = 7.5); *Education:* not reported Sex not reported	Short orientation-memory-concentration test; star cancellation task	Not reported	Participants were included if they could perform the word fluency task (part of dual-task protocol)
Kim et al. (2021)	10 stroke; 7 controls	*Mean age* = 54.7 (SD = 12.3); *Education*: not reported *Sex* = all M	MMSE	Not reported	Participants with communication disorders due to aphasia excluded
Laganaro et al. (2019)	12 stroke; 11 neurotypical controls	*Stroke: mean age* = 69.1 (SD = 13.2), 6 M, 6 F; *Controls: mean age* = 63.1 (SD = 14.6); *Sex* = 4 M, 7 F	Language (naming, repetition, spoken comprehension); Go/No Go task	Not reported	Mild aphasia (with mild anomia, and some lexical and/or phonological paraphasias); other language functions also tested
Lee et al. (2021)	13 stroke	*Mean age* = 45.9 (SD = 11.9) *Mean education* = 12.4 years (SD = 4.4.) *Sex* = all M	MMSE	Not reported	Aphasia was an exclusionary criterion but not otherwise defined or assessed
Murray (2000)	14 LH stroke; 7 RH stroke; 1 RH glioblastoma; 9 neurotypical controls	*Stroke LH: mean age* = 64.1 (SD = 11.4); *Mean education* = 13.9 years (SD = 3.4); *Sex* = 9 M, 5 F; *Stroke RH: mean age* = 71.8 (SD = 10.8);	Estimated IQ; speech discrimination; ADP in both clinical groups; Limb apraxia test (aphasia group only); BIT; MIRBI (RH group only)	Not reported	Aphasia as per ADP (LH group only); RH group assessed with MIRBI

		Mean education = 12.5 years (SD = 2.0); *Sex* = 5 M, 3 F; *Controls: mean age* = 62.8 (SD = 13.3); *Mean education* = 15.4 years (SD = 2.8); *Sex* = 6 M, 3 F			
Tsang et al. (2019)	30 stroke	*Mean age* = 62.4 (SD = 6.7); *Highest education level* = 11 primary, 16 secondary, 3 tertiary; *Sex* = 22 M, 2 F;	MoCA	Not reported	Participants were able to follow given instructions
Tsang & Pang (2020)	91 stroke	*Mean age* = 62.7 (SD = 8.3); *Mean education* = 9.4 years (SD = 4.3); *Sex* = 64 M, 27 F	MoCA; Stroop	Not reported	Participants were able to follow given instructions
Yang et al. (2016)	88 participants	*Mean age* = 62.9 (SD = 7.8), range = 50 – 80; *Mean education* = 9.3 years; range = 0 – 19 years *Sex = 64 M,* 24 F	MoCA	Not reported	Participants were included if they could perform two-stage commands

Notes. ADP - Aphasia Diagnostic Profiles (Helm-Estabrooks, 1992); BIT - Behavioural Inattention Test (Cockburn et al., 1987); MIRBI - Mini Inventory of Right Brain Injury (Pimental & Kingsbury, 1989); MoCA – Montreal Cognitive Assessment (Nasreddine et al., 2005); MIRBI - Mini Inventory of Right Brain Injury (MIRBI; Pimental & Kingsbury, 1989); MMSE - Mini Mental State Examination (Folstein et al., 1975); NART - National Adult Reading Test (Nelson, 1982); bedside WAB - Western Aphasia Battery (Shewan & Kertesz, 1980); WAIS - Wechsler Adult Intelligence Scale (citation not reported).

Table 3 Single word studies – Stroke information

Studies	Stroke type	Lesion site	Stroke severity	Time post-onset
Bhatt et al. (2016)	5 ischaemic; 5 haemorrhagic	6 left-sided; 4 right-sided	Not reported	*Mean* = 8.93 years (SD = 3.07)
Bruehl et al. (2021)	Not reported	Left-sided in all 19 participants	Not reported	*Mean = 26 months, range* = 4 - 63
Cockburn et al. (2003)	6 ischaemic; 2 haemorrhagic; No information for 2 participants	2 left-sided; 8 right-sided	Not reported	*Mean* = 5.7 months, range = 1 - 10
Feld & Plummer (2021)	36 ischaemic; 11 haemorrhagic	12 right-sided; 8 left-sided; 19 subcortical; 5 bilateral; 3 brainstem	*Median* NIHSS score on admission (of 34 participants) = 6 (IQR = 3 – 9)	*Median* = 14 days (IQR = 7 – 21)
Haggard et al. (2000)	30 cortical stroke; 3 brainstem; 5 subarachnoid haemorrhage	22 right-sided; 8 left-sided	Not reported	*Mean* = 16.36 months (SD = 30.77), *range* = 1 - 156
Harley et al. (2006)	Not reported	Not reported	Not reported	*Mean* = 69 days (SD = 50)
Kim et al. (2021)	Not reported	Not reported	Not reported	*Mean* = 50.9 months
Laganaro et al. (2019)	Not reported	Left-sided in all 12 participants	Not reported	*Range* = 0.5 – 97 months
Lee et al. (2021)	Ischaemic in all 13 participants	Right-sided in all 13 participants	Not reported	Chronic phase of stroke but not defined
Murray (2000)	Not reported	Not reported	Not reported	*LH mean* = 33.3 months (SD = 36);

				RH mean = 10.0 months (SD = 4.5)
Tsang et al. (2019)	19 ischaemic stroke; 9 haemorrhagic; 2 unknown	Not reported	Not reported	*Mean* = 9.2 years (SD = 3.6)
Tsang & Pang (2020)	Not reported	Not reported	Not reported	*Mean* = 8.8 years (SD = 5.3)
Yang et al. (2016)	Not reported	Not reported	Not reported	Mean = 105.9 months, SD = 61.6 (Table 1)

Notes. CMSA - Chedoke-McMaster Stroke Assessment (Gowland et al., 1993);

GDS - Geriatric Depression Scale (Short Form) (Wong et al., 2002); Fugl Meyer (citation not provided);

NIHSS - National Institutes of Health Stroke Scale (Brott et al., 1989).

Table 4 Single word studies – Task prioritisation, reliability

Studies	Task prioritisation	Reliability / fidelity information
Bhatt et al. (2016)	No instructions to prioritise either task.	Not reported.
Bruehl et al. (2021)	Not reported.	Not reported.
Cockburn et al. (2003)	No instructions to prioritise either task.	Not reported.
Feld & Plummer (2021)	No instruction to prioritise either task.	Not reported.
Haggard et al. (2000)	No instructions to prioritise either task but asked to attempt to combine both tasks.	Not reported.
Harley et al. (2006)	Not reported.	Not reported.
Kim et al. (2021)	Not reported.	Not reported.
Laganaro et al. (2019)	Not reported.	Inter-rater reliability on error coding; discrepancies were resolved by discussion.
Lee et al. (2021)	Participants asked to focus on motor task.	Not reported.
Murray (2000)	Participants asked to respond first to the tone discrimination task and second to the phrase completion.	*Inter-rater agreement:* The responses of three aphasic, two RBD, and two NBD subjects (i.e., approximately 20% of the total responses) were randomly selected for re-transcription by a second listener blind to subject group. Agreement for words transcribed was 99%. *Intra-rater agreement:* 20% of the responses (i.e., 3 LH, 2 RH, 2 controls) randomly selected and re-analysed by the author more than a month after original scoring and coding. Intra-judge agreement was 100% for accuracy and 96.5% for errors.
Tsang et al. (2019)	Participants asked to perform both tasks equally well.	Test-retest reliability carried out within 7 to 14 days after initial assessment; purpose of study was to assess test re-test reliability; fidelity information about coding of participants' responses was not reported.

Tsang & Pang (2020)	Not reported.	Not reported.
Yang et al. (2016)	Participants asked to perform both tasks equally well.	Test-retest reliability carried out 3-4 days after initial assessment; fidelity information about coding of participants' responses was not reported.

Table 5 Single word studies – task type and single vs. dual-task comparisons, results of comparisons

Studies	Task type and single vs. dual-task comparisons	Results of dual-task effects on spoken language output
Bhatt et al. (2016)	Letter fluency (lasting 8 seconds) while sitting vs. carrying out a voluntary balance-controlled task (forward direction with harness support).	Statistically significant dual-task costs between stroke group in comparison to the other two neurotypical groups (older and younger adults).
Bruehl et al. (2021)	Spoken picture naming (without interference) vs. interfered naming with five auditory distraction conditions: i) phonological; ii) associative semantic; iii) categorical semantic; iv) unrelated condition; v) noise distraction.	Statistically worse performance ($p < .1$; alpha level set by the authors) in 11 of the 19 participants between spoken picture naming vs. interfered naming.
Cockburn et al. (2003)	Word fluency ("things to eat") while sitting vs. while walking.	Descriptive statistics reported individually; across 5 participants there were dual-task costs ranging from 7% to 57%; across the other 5 participants there was a dual-task gain ranging from 12% to 39%. At group level, a related samples Wilcoxon test between single- and-dual-task performance (number of unique words produced) carried out by the present review's authors was not statistically significant, $z = .82$, $p = .414$ (two-tailed).
Feld & Plummer (2021)	Word fluency while sitting vs. word fluency score while walking for one minute at self-selected speed. Word fluency tasks, either easy (e.g., items of furniture/clothing, occupations) or difficult (e.g., appliances/other electronics, tools); participants were assigned either to "easy" or "difficult" based on their performance on an animal fluency task (as a single-task).	No statistically significant difference between single- and dual-task performances ($p = .078$).
Haggard et al. (2000)	Word fluency while sitting vs. word fluency while walking. Semantic fluency tasks (one from the following: "things to eat", "things to drink", "things in the house" or "things in the street").	Authors carried out within group comparisons separately for stroke and control groups. Dual-task effects (costs) were found in the stroke group. Descriptive statistics of z scores (mean [SD]) for dual-task scores between stroke and control groups: 14.87 (5.36) vs. 24.50 (4.30) [CI = 13.4 – 16.4] vs. [CI = 21.8 – 27.2]

Harley et al. (2006)	Word fluency while supported vs. unsupported sitting. Word fluency was in one of four semantic categories (as in Haggard et al., 2000).	Authors carried out within group comparisons separately for stroke and control groups. Dual-task effects (costs) were found in the stroke group. Descriptive statistics (mean (SD)) for dual-task scores between stroke and control groups: 13.8 (6.4) vs. 24.1 (5.9) [CI = 11.7 – 15.9] vs. [CI = 21.6 – 26.2]
Kim et al. (2021)	Word fluency vs. word fluency while performing cognitive upper limb motor tasks (one involving circle shaped movement and one with a cross shaped movement). Word fluency was either letter fluency (words beginning with "B") or semantic fluency (hospital-related words).	No dual-task effect differences within or between groups.
Laganaro et al. (2019)	Picture naming accuracy vs. picture naming accuracy under two conditions in: i) passive dual-task of ignoring syllables (verbal) or tones (non-verbal); ii) active dual-task of detecting auditory stimuli by pressing a button while naming.	Statistically significant dual-task costs were found in stroke (people with aphasia): mean accuracy in single-task naming = 69.4%, passive dual-task naming = 69%, active dual-task naming 65.6%
Murray (2000)	Single word production elicited by phrase completions that had a relatively closed set of responses (e.g., ''he milked the …'') or an open set of responses (e.g., "he carried the …'') vs. word production while also completing a tone discrimination task.	In dual-task, word retrieval accuracy was significantly less than in single-task in both stroke groups (left-hemisphere stroke or aphasia and right-hemisphere stroke). Phrase type only affected the left-hemisphere stroke group. Controls were not affected by any manipulations (dual-task or phrase type).
Lee et al. (2021)	Each word fluency task (animals, supermarket items, phonemic – Korean letters) as single-task vs. dual-task with an upper limb motor task.	Descriptive statistics of z scores (mean (SD)) between single- vs. dual-task (5th, baseline measurement [1] and 95% CIs): Animal: - 1.309 (SD = .985) vs. - 1.407 (SD = .949) [CI = - 1.84 to - .774] vs. [CI = - 1.92 to - .891] Supermarket items: - 1.282 (SD = .795) vs. - 1.682 (SD = .728) [CI = -1.71 to - .85] vs. [CI = -2.08 to -1.29] Phonemic: - .422 (SD = 1.247) vs. - .765 (SD = 1.212)

		[CI = - 1.1 to .256] vs. [CI = -1.42 to - .106]
Tsang et al. (2019)	Word fluency tasks with two levels of difficulty: i) low – words from a randomly selected category (e.g., countries); ii) high – words from a more confined randomly selected category (e.g., European countries). Fluency tasks were presented while walking on level ground (simple dual-task) vs. walking with obstacle crossing (complex dual-task).	Based on descriptive statistics (means, SDs, table 3), the present review's authors calculated 95% CI with results as follows: Low difficulty: simple dual-task - *mean* = 12.8 (SD = 5.2) [CI = 11.0 to 14.6] complex dual-task - *mean* = 13.1 (SD = 4.3) [CI = 11.6 to 14.6] High difficulty: simple dual-task – *mean* = 8.3 (SD = 4.7) [CI = 6.6 to 10.0] complex dual-task task – mean = 8.0 (SD = 3.1) [CI = 7.0 to 9.1]
Tsang & Pang (2020)	Word fluency in isolation was compared to word fluency while: i) forward walking (simple dual-task); ii) obstacle-crossing (complex dual-task).	Based on descriptive statistics (means, SDs, table 3), the present review's authors calculated 95% CIs as follows: Simple dual-task – mean = .5 (SD = .2) [CI = 0.46 to 0.54] Complex dual-task – mean = .4 (SD = .2) [CI = 0.36 to 0.44]
Yang et al. (2016)	Word fluency ("fruits") while walking at comfortable speed (simple dual-task) vs. word fluency ("clothes naming") while completing a timed up and go task (complex dual-task).	Based on descriptive statistics (means, SDs, table 3), the present review's authors calculated 95% CIs as follows: Simple dual-task – mean = .47 (SD = .18) [CI = .432 to .508] Complex dual-task – mean = .34 (SD = .14) [CI = .311 to .369]

Notes. [1] as described in Method section, only the first, baseline measurement reported; data reported here were extracted from Tables 1 and 2 (Lee et al., 2021).

Table 6 Methodological appraisal of single word studies

	Rating categories				
Studies	Study design	Demographic variables	Stroke variables	Cognitive variables	Overall rating
Bhatt et al. (2016)	moderate	moderate	moderate	low	low
Bruehl et al. (2021)	moderate	high	moderate	low	low
Cockburn et al. (2003)	low	moderate	low	moderate	low
Harmon et al. (2019)	high	high	low	moderate	low
Feld & Plummer (2021)	moderate	high	moderate	moderate	moderate
Haggard et al. (2000)	high	moderate	low	low	low
Harley et al. (2006)	moderate	low	low	moderate	low
Kim et al. (2021)	moderate	moderate	low	low	low
Laganaro et al. (2019)	high	high	low	moderate	low
Lee et al. (2012)	moderate	high	moderate	low	low
Murray (2000)	high	high	low	low	low
Tsang et al. (2019)	moderate	high	moderate	low	low
Tsang & Pang (2020)	moderate	high	low	moderate	low

Yang et al. (2016)	moderate	moderate	low	low	low

Note. Overall rating: For high rating, a study must score High in 3/4 categories (with no Low rating); For a moderate rating, a study cannot receive any low rating.

Table 7 Discourse studies – Participant information

Studies	Participants	Demographics	Cognitive-linguistic variables	Motor speech disorder information	Information about aphasia
Harmon et al. (2019)	19 stroke; 1 brain injury; 1 multiple sclerosis; 12 neurotypical controls	*Stroke: mean age* = 59 (range = 32 – 81) *Mean education* = 16 years (range 12 -22) *Sex* = 5 M, 13 F *Controls: mean age* = 58 (range = 33 – 81), *Mean education* = 16 years (range 12 – 21) *Sex* = 5 M, 7 F	TALSA TONI-4 CCRSA	10 participants had apraxia of speech and/or dysarthria	All participants presented with aphasia according to the WAB-R of either mild or moderate severity based on authors' severity criteria
Kemper et al. (2006)	10 stroke; 10 neurotypical controls	*Stroke: mean age* = 77.2 (SD = 5.8); *Controls: mean age* = 76.3 (SD = 5.4); *Sex:* groups were matched but no other information reported	PCSC (Short); Shipley vocabulary test; WAIS (digit span - forward, backward); digit symbol substitution test); Stroop test	Not reported	No aphasia according to ADP
Murray et al. (1998)	14 stroke; 8 controls	*Stroke: mean age* = 64.07 (SD = 11.43), *Mean education* = 13.9 (SD = 3.36) *Sex* = 10 M, 4 F; *Controls: mean age* = 62.50 (SD = 14.20) *Mean education* = 15 (SD = 2.62) *Sex* = 5 M, 3 F	Limb apraxia (ABA-2); Estimated IQ	Not reported	All participants presented with aphasia according to ADP

Plummer-D' Amato et al. (2008))	13 stroke	*Mean age* = 60.5 (SD = 15.3); *Education* not reported; *Sex* = 11 M, 2 F	MMSE; DST; WMS (backward digit span); WASI (vocabulary test); Digit ordering; Stroop	No severe dysarthria	Participants who had severe aphasia and/or could not follow three-step commands were excluded
Plummer et al. (2020)	29 stroke; 23 controls	*Stroke: Mean age* = 54.2 (SD = 16); *Sex* = 15 M, 14 F; *Mean education* = 14.9 years (SD = 2.9) *Controls: Mean age* = 54.5 (SD = 15.3); *Sex* = 10 M, 13 F; *Mean education* = 16.2 years (SD = 2.9)	MoCA; WAIS (vocabulary); RBANS; WAB	Not reported	Participants with stroke could not have severe aphasia, assessed with bedside WAB; 6 (21 %) participants with Aphasia Quotient score below 93.8, (mild, anomic aphasia)
Pohl et al. (2011a)	19 stroke	*Mean age* = 66.8 (SD = 8.4); *Education* not reported; *Sex* = 13 M, 6 F	MMSE	Not reported	Absence of aphasia determined by lexical retrieval tests of the ADP
Pohl et al. (2011b)	24 stroke; 12 neurotypical controls	*Stroke: Mean age* = 66.5 (SD = 9.1); *Mean education* = 15.2 years (SD = 2.9) *Sex* = 33% F, 67% M; *Controls: Mean age* = 72.7 (SD= 8.0); *Mean education* = 17.1 years (SD = 1.6) *Sex* = 50% F, 50& M;	MMSE; WAIS (digit span -forward, backward; digit symbol); Shipley vocabulary	Not reported	Absence of aphasia was determined by lexical retrieval tests of the ADP
Rogalski et al. (2010)	13 stroke	*Mean age* = 60.46 (range = 33 – 86); *Education* not reported *Sex* = 11 M, 2 F	MMSE; Stroop; WMS (digit symbol, backward digit span); digit ordering test; WASI vocabulary	No severe dysarthria	Participants screened by certified speech-language pathologists; none met clinical criteria for aphasia

Notes. TALSA - TALSA Temple Assessment of Language and Short-term memory in Aphasia (Martin et al., 2012); TONI-4 - Test of Nonverbal Intelligence (4th ed.) (citation not provided); WAB-R - Western Aphasia Battery (Kertesz, 2007); CCRSA - Communication Confidence Rating Scale for Aphasia (Babbitt et al., 2011); PCSC (Short) - Portable Cognitive Status Questionnaire (Pfeiffer, 1975); ADP - Aphasia Diagnostic Profiles (Helm-Estabrooks, 1992, p.); WAIS – Wechsler Adult Intelligence Scale (Wechsler, 1958); WASI - Wechsler Abbreviated Scale of Intelligence (Wecshler, 1999); MMSE; DST; WMS (backward digit span); WASI (vocabulary test); Digit ordering (Hoppe et al., 2000); Stroop (Stroop, 1935).

Table 8 Discourse studies - Stroke information

Studies	Stroke type information	Lesion site	Stroke severity	Time post-onset
Harmon et al. (2019)	Not reported	Not reported	Not reported	> 19 months
Kemper et al. (2006)	Not reported	3 RH; 5 LH; 2 bilateral	Not reported	*Range* = 24 – 36 months
Murray et al. (1998)	Not reported	LH in all (5 frontal, 6 temporo-parietal, 1 basal ganglia, 1 temporal, 1 parietal)	Not reported	*Mean* = 33.29 months (range = 6 – 75)
Plummer-D' Amato et al. (2008)	Ischaemic in 12 participants; not reported for 1 participant	Not reported	Not reported	*Mean* = 8.7 months (SD = 4.8)
Plummer et al. (2020)	20 ischaemic stroke; 9 haemorrhagic;	Not reported	SIS (self-perceived impact)	*Median* = 10.7 months, range = 1.5 – 36.3
Pohl et al. (2011a)	Not reported	unclear	Not reported	Mean = 49.5 months (SD = 34.7)
Pohl et al. (2011b)	Not reported	Not reported	SIS (recovery)	*Mean* = 46.3 months (SD = 32.3)
Rogalski et al. (2010)	12 ischaemic; 1 not reported	Not reported	Not reported	*Mean* = 8.69 months (SD = 4.79)

Notes. WAIS - Wechsler Adult Intelligence Scale (citation not reported); Fugl Meyer (citation not provided);

SIS - Stroke Impact Scale.

Table 9 Discourse studies – Task prioritisation and reliability

Studies	Task prioritisation	Reliability / fidelity information
Harmon et al. (2019)	Not reported	*Intra-rater:* All reliability measures were obtained by having the primary coder(s) recode a random selection of 20% of the language samples (word count, utterance count, CIUs, repetition, pauses). *Inter-rater:* Also based on a random selection of 20% of the language samples (word count, utterance count, CIUs, repetition, pauses). Range of correlations for inter- and intra-rater was .80 - .99
Kemper et al. (2006)	Participants asked to respond to question without interrupting the concurrent activity	Two trained coders independently scored 10% of the language samples; agreement exceeded, r (15) > .90 for all measures
Murray et al. (1998)	Participants instructed to give priority to picture description and guess in the listening task	*Inter-rater:* 24 transcripts (i.e., 4 aphasia and 2 controls 4 speaking conditions) randomly selected, with rater blinded to subject group or condition identity. Point-to-point inter-judge agreement across measures ranged from 93% to 100%. *Intra-rater:* 12 transcripts (2 aphasia and 1 control 4 speaking conditions) randomly selected and re-analyzed by the first author at least one month after the original linguistic and pragmatic coding. Point-to-point intra-judge agreement across measures ranged from 86% to 100%.
Plummer-D' Amato et al. (2008)	Participants not specifically instructed to prioritise either task	Not reported
Plummer et al. (2020)	*Laboratory setting*: talking vs. talking while prioritising walking *Real world setting:* talking vs. talking while prioritising walking	Not reported
Pohl et al. (2011 a)	Participants not given any instructions to prioritise one task or the other	Not reported

Pohl et al. (2011 b)	Participants not given any instructions to prioritise one task or the other	20% of audio recordings were analysed independently by a research assistant. For all language measures agreement was r > .90
Rogalski et al. (2010)	Participants not instructed to prioritise either walking or talking	*Intra-rater reliability:* a random 20% of transcripts were re-transcribed about 3 months after initial coding. Results: 88.49% for global coherence; 91.18% for local coherence. *Inter-rater reliability:* a research assistant trained on a coherence rating scale. Results: 85.09% for global coherence; 87.47% for local coherence.

Note. CIUs – Correct Information Units

Table 10 Discourse studies – Results

Studies	Single-dual task comparisons	Results of dual-task effects on spoken language output
Harmon et al. (2019)	Story retelling as single-task vs. while simultaneously discriminating auditorily presented (high or low frequency tones) by pressing corresponding buttons. *Measures:* number of correct information units (CIUs: intelligible, accurate, relevant, and informative words), efficiency (ratio of correct CIUs per word), speech rate (number of words per minute), pauses (ratio of extended and filled pauses per utterance).	*Moderate aphasia:* statistically significant dual-task costs in: number of correct CIUs, ratio of CIUs per word, speech rate; no dual-task costs in pauses per utterance. Performance in number of correct CIUs and ratio of CIUs per word was significantly lower than controls; no difference with controls in speech rate or pauses per utterance. *Mild aphasia:* statistically significant dual-task costs in number of correct CIUs, speech rate and pauses per utterance, but not in ratio of CIUs per word. In comparison to controls only speech rate differed (slower).
Kemper et al.	A talking alone condition vs.	*Dual-task costs found in the following:*

(2006)	i) talking while ignoring speech; ii) talking while ignoring noise; iii) walking while talking; iv) simple finger tapping while talking; v) complex finger tapping while talking. Language samples elicited from questions asking participants to describe people or events that have influenced their lives, recent vacations, inventions of the 20th century, individuals they admire. *Measures:* Number of words per minute; % of grammatical utterances; mean length of utterance; syntactic complexity; type-token ratio; propositional density	*Speech rate:* talking while ignoring noise, speech, walking, simple, and complex tapping. *% of grammatical utterances:* talking while ignoring noise (but not speech), walking simple, and complex tapping. Group x task interaction not significant. *Mean length of utterance:* talking while ignoring noise, speech, walking, simple, and complex tapping. *Syntactic complexity:* talking while ignoring noise, speech, walking, simple, and complex tapping. *Type-token ratio:* talking while ignoring noise, speech, simple, and complex tapping (but not while walking). *Propositional density:* talking while ignoring noise, speech, walking, simple, and complex tapping.
Murray et al. (1998)	Picture description alone vs. in competition with a tone-discrimination task: i) picture and tones presented simultaneously; participants required to complete both tasks and guess in the tone, if necessary; ii) picture and tones presented simultaneously but required to complete both tasks. *Morphosyntactic measures:* grammatical and ungrammatical simple and complex sentences; verb morphology. *Lexical and pragmatic measures:* Correct information units (CIUs), unsuccessful utterances	Stroke (aphasia group) showed a significant decrease in the proportion of syntactically complete utterances in dual-task ii condition as compared to single task. In both dual-task conditions, they also produced more simple sentences than controls. Number of simple sentences in controls did not significantly differ by task manipulation. Verb morphology was not affected by task manipulation in either group. *Lexical and pragmatic measures:* significant decrease from single- to dual-tasks (both) in the stroke group and a marginally significant effect (p = .062) between the two dual-tasks. CIUs and unsuccessful utterances: significant reductions between single- vs. dual-task conditions (both) in the stroke group. No dual-task effects in controls.
Plummer-D' Amato et al. (2008)	Talking while sitting vs. while walking (oval track). Language samples were elicited by asking participants to describe people or events that have influenced their lives, recent vacations, 20^{th} century inventions, individuals they admire.	Significant dual-task effects with decrease in number of utterances and words per narrative and increase in number of pauses per utterance and proportion of utterances with new information; no dual-task effects in sentence length, sentence complexity, proportion of grammatical sentences.

	Measures: sentence length (words per utterance), number of utterances, number of words, number of fillers per utterance, number of pauses per utterance, sentence complexity, proportion of grammatical sentences, proportion of utterance with new information	
Plummer et al. (2020)	Talking vs. talking while walking in lab hallway: Talking vs. while walking in hospital lobby Language samples elicited with questions where participants talked about previous experiences (e.g., favourite holiday) or opinions (e.g., most important invention in the last century). *Measures:* proportion of sentences with new information, number of verbs per sentence.	*Environment speech-gait dual-task:* Stroke group produced fewer sentences with new information than controls in both single- and dual-task (no dual-task cost). Stroke group produced significantly fewer verbs per sentence than controls (dual-task cost). *Priority instruction speech-gait dual-task:* No significant main effects or interactions on dual-task effects for speech variables.
Pohl et al. (2011a)	Talking alone vs. talking while a) moving the affected hand; b) moving the less affected hand. Language samples elicited with questions about participants' experiences in their lives or opinions, such as describing a favourite holiday or a significant invention in the 20th century. *Measures:* Speech rate (words per minute) which included words and lexical fillers (e.g., you know, well) but excluded filled pauses (e.g., uh, duh).	Speech rate while moving the affected hand was statistically higher than speech rate in the single-task. Speech rate while moving the less affected hand was not significant (p = .068).
Pohl et al. (2011b)	Talking alone vs. talking while walking. Language samples elicited with questions asking participants about their experiences or	Dual-task costs statistically greater in the stroke group vs. controls in terms of mean clauses per utterance and syntactic complexity. No group differences in speech rate, MLU, % f grammatical utterances, propositional density, and type-token ratio.

	opinions, such as describing a favourite holiday, a person they admire. *Measures:* Speech rate (words per minute) which included words and lexical fillers (e.g., you know, well) but excluded filled pauses (e.g., uh, duh); mean length of utterance (MLU); % of grammatical sentences; mean clauses per utterance; syntactic complexity; propositional density; type-token ratio	
Rogalski et al. (2010)	Talking while sitting vs. talking while walking. Participants spoke for at least 2 minutes on topics including people or events that had had a significant impact on their lives, as well as ‘‘Tell me what you like or dislike about the city you grew up in?’’ or ‘‘What sort of things would you consider important in a marriage partner?” *Measures:* Discourse coherence (global, local)	No significant main effect for task (single-dual), but significant main effect of type of coherence with global coherence scores significantly lower than local coherence scores, regardless of task. No interaction between task and coherence type.

Table 11 Methodological appraisal of discourse studies

	Rating categories				
Studies	Study design	Demographic variables	Stroke variables	Cognitive variables	Overall rating
Harmon et al. (2019)	high	high	low	moderate	low
Murray et al. (1998)	high	high	high	low	low
Kemper et al. (2006) (	moderate	moderate	low	high	low
Plummer-D'Amato et al. (2008)	moderate	moderate	low	high	low
Plummer et al. (2020)	high	high	moderate	high	high
Pohl et al. (2011a)	moderate	moderate	low	low	low
Pohl et al. (2011b)	high	high	low	high	low
Rogalski et al. (2010)	moderate	moderate	low	high	low

Note. Overall rating: For high rating, a study must score High in 3/4 categories (with no Low rating); For a moderate rating, a study cannot receive any low rating.