Single-Word and Conversational Measures of Word-Finding Proficiency

Susan J. Tingley*
Christiane S. Kyte
Carla J. Johnson
University of Toronto, Ontario, Canada

Joseph H. Beitchman
Child and Family Studies Centre, Centre for Addiction and Mental Health, Clarke Division, Toronto, Ontario, Canada

Two studies with young adults as participants evaluated the relationship, presumed in the word-finding literature to exist, between slow, inaccurate performances in single-word-naming and semantic-retrieval tasks and disruptions to conversational fluency. The measures evaluated were the frequency of conversational disruptions and the scores from 3 single-word tasks: total time from the Rapid Automatized Naming task (RAN; M. B. Denckla & R. G. Rudel, 1976), standard score from the Brief Test of the Test of Adolescent/Adult Word Finding (TAWF; D. J. German, 1990), and total unique words from the Controlled Oral Word Association task (FAS; A. L. Benton & K. Hamsher, 1978). RAN time was the only significant predictor of the frequency of conversational disruptions, although this relationship was weak (R^2 = .11). In addition, single-word performances did not discriminate between groups of participants with differing levels of conversational fluency. Clinicians are cautioned against identifying word-finding deficits using single-word measures alone. Moreover, the theoretical construct of word-finding difficulties requires additional validation.

Key Words: word finding, conversational fluency, naming, word retrieval, language assessment

Word-finding difficulties are thought to be characteristic of developmental and acquired language disorders (e.g., Goodglass, Wingfield, & Hyde, 1998; McGregor & Leonard, 1995). A word-finding impairment is typically defined as a reduced ability to recall and/or produce a specific word in response to a stimulus or situation (Faust, Dimitrovsky, & Davidi, 1997; Hall & Jordan, 1987; Rapin & Wilson, 1978). Such difficulties presumably impede oral communication and reading, both of which require efficient retrieval of words from the mental lexicon (Faust et al., 1997; Snyder & Godley, 1992). Accurate identification of word-finding problems is, therefore, an important undertaking for the speech-language pathologist. Word-finding problems may be evident in the production of single words, connected discourse, or both (German, 1992).

A variety of single-word tasks are thought to tap word-finding abilities (see Snyder & Godley, 1992, for a review). These tasks generally involve naming in response to visual or auditory stimuli (e.g., pictures, letters, definitions, rhyme cues) or retrieval of information from semantic memory (e.g., animal names, words beginning with *s*-). The common element across the various tasks is the requirement for production of a specific word or series of words. Accuracy and/or speed of production are typically measured, with slow, inaccurate performances thought to indicate word-finding difficulties.

There is ample and consistent evidence that individuals with language disorders show reduced speed and accuracy of responses on single-word tasks such as naming and semantic retrieval relative to peers without language disorders. Such evidence is available for both acquired language disorders (Goodglass & Kaplan, 1972; Goodglass et al., 1998; Knopman, Selnes, Niccum, & Rubens, 1984) and developmental language disorders (Denckla & Rudel,

* Currently affiliated with University Health Network, Toronto, Ontario, Canada.

1976; Faust et al., 1997; Kail & Leonard, 1986; Lahey & Edwards, 1996; Wiig & Semel, 1975). The diagnosis of word-finding difficulties based on poor performance on single-word naming and retrieval measures is not, however, straightforward because such difficulties may in part reflect inadequate storage of lexical items, rather than a difficulty related uniquely to the speed and accuracy of word retrieval (Kail & Leonard, 1986). Possible subtle differences in the strength and elaboration of items in lexical storage are difficult to rule out definitively, even when there is clear evidence of comprehension for single words that are not retrieved (McGregor & Leonard, 1995).

Connected speech may also contain symptoms of word-finding problems, in the form of disruptions or breakdowns in the fluency of language formulation (Faust et al., 1997; Nippold, 1992; Snyder & Godley, 1992). These conversational disruptions may reflect incomplete word knowledge, erroneous retrieval processes, or stalling tactics that allow a speaker more time to generate the intended word (Hall & Jordan, 1987; MacLachlan & Chapman, 1988; Wiig & Semel, 1984). Disruptions proposed to index word-finding problems include the use of nonspecific, empty words (e.g., *stuff*, *thing*), filled pauses (e.g., *um*, *er*), silent pauses, circumlocutions (e.g., *thing to open doors*), metalinguistic comments on language formulation (e.g., *I can't think of the word*), word substitutions (e.g., *dog/cat; jogging/juggling*), utterance fillers (e.g., *I mean*, *whatever*), phrase repetitions, and statement reformulations (German, 1987, 1992; German & Simon, 1991; Snyder & Godley, 1992).

The limited existing evidence is unclear on whether individuals with language disorders exhibit disruptions in conversational fluency more frequently than those without disorders (Dollaghan & Campbell, 1992; Jordan, Ward, & Cremona-Meteyard, 1997; MacLachlan & Chapman, 1988; Scott & Windsor, 2000; Wiig & Semel, 1975). Group differences, when observed, have not been either pervasive or consistent across various types of conversational disruptions. Further, it is not clear that frequent conversational disruptions specifically reflect word-finding difficulties, rather than stuttering or the influences of other factors, such as the syntactic complexity of the discourse required (MacLachlan & Chapman, 1988) or the speaker's knowledge of the topic being discussed (Perry & Lewis, 1999). Given these ambiguities, it seems prudent to assess further the presumed underlying nature of word-finding difficulties.

In this research, we assess the construct of word-finding using both single-word and connected speech measures. We test an assumption implicit in much of the word-finding literature, namely, that lower accuracy scores and slower response times on single-word tasks are associated with higher frequencies of disruptions in connected speech. This key assumption enjoys surprisingly little empirical support.

German and colleague (German, 1987; German & Simon, 1991) investigated the word-finding abilities of school-aged children in single-word tasks and connected speech. Children with word-finding difficulties were identified based on a variety of criteria, including poor performances on various single-word tasks, enrollment in language remediation involving word-finding goals, and identification of presumed word-finding symptoms (including conversational disruptions) using a subjective checklist completed by a speech-language pathologist. This group was then compared to children without word-finding difficulties with respect to observed frequencies of conversational disruptions (e.g., substitutions, reformulations, repetitions, silent/filled pauses, empty words) in picture description tasks. Children with word-finding problems produced significantly more disruptions per utterance than their peers without such difficulties.

Unfortunately, definitive conclusions regarding the relationship between single-word and connected speech measures cannot be made based on these findings (German, 1987; German & Simon, 1991). The ambiguity in interpretation arises because connected speech measures were compared for groups that were identified based on a combination of single-word measures and the subjective checklist measure, rather than single-word measures alone. The checklist required subjective judgments of the frequencies of conversational disruptions (i.e., items assessing substitutions, reformulations, empty words, metalinguistic comments, filled pauses; see German, 1983). The findings may therefore show that those who were subjectively rated as experiencing poor conversational fluency showed more frequent disruptions than did their peers when these behaviors were measured objectively. Thus, the findings may demonstrate a relationship between subjective and objective measures of conversational fluency, rather than a relationship between single-word measures and conversational fluency.

Jordan et al. (1997) investigated the word-finding abilities of children with and without a history of severe closed head injury (CHI) in single-word naming tasks and in conversation. Accuracy scores for picture naming were compared with the frequencies of conversational disruptions per 100 words during a guided interview. Disruptions analyzed included repetitions, revisions, orphans, and pauses (silent, filled, pause strings). Children with and without CHI differed significantly on the frequencies of silent pauses and pause strings, but not on other disruption types. Naming scores did not predict the frequency of pauses in children with CHI. However, a strong conclusion regarding the usefulness of single-word measures in predicting conversational fluency was prevented because many types of disruptions were excluded from the correlational analysis (e.g., repetitions, revisions).

Heller and Dobbs (1993) evaluated word-finding proficiency in connected speech (video description) and speeded semantic retrieval in normally aging adults. Multiple regression analyses assessed whether semantic retrieval performance predicted several types of conversational disruptions. The use of nonspecific object labels was the only disruption type predicted by retrieval performance, and the predictive relationship was weak (ΔR^2 = .108).

Taken together, the existing findings do not provide strong support for the presumed relationship between single-word measures and conversational fluency. Accordingly, we conducted two exploratory studies to further

address this issue. The first used a correlational design to evaluate the strength of the relationship between performance on single-word measures and the frequency of conversational disruptions in young adults, with and without language disorders, who demonstrated a wide range of language abilities. If single-word proficiency and conversational fluency are related to each other via the construct of word-finding ability, a correlational analysis across a broad range of abilities should expose the relationship.

The second study used a group design to examine whether young adults subjectively rated as having poorer conversational fluency demonstrated lower single-word scores than those rated as having adequate conversational fluency. The two studies were intended to provide converging evidence on possible relationships between single-word performances and conversational fluency.

General Method

The two studies reported here were based on secondary analyses of data originally collected for other purposes from young adults in the Ottawa Language Study (OLS). The OLS is an ongoing, prospective, longitudinal investigation of the natural history of 284 children (142 with communication disorders and 142 matched control children). The OLS began in 1982 when the participants, then 5 years old, underwent comprehensive speech-language, cognitive, and psychosocial assessments (Beitchman, Nair, Clegg, & Patel, 1986). Two similar follow-up assessments of the OLS sample occurred at ages 12 and 19. At age nineteen, 242 young adults from the original sample of 284 received complete speech-language assessments (Johnson, Beitchman, et al., 1999).

Participants

For the current studies, we selected participants from those who received speech-language assessments at age 19 in the OLS. Individuals who stuttered ($n = 5$) were excluded to avoid a possible confounding of stuttering with other types of conversational disruptions. The current participants were therefore drawn from the remaining pool of 237 eligible young adults. Selection criteria for each study are described later in this report.

Criteria for Language Impairment

At age 19, OLS participants were considered to have a language impairment if they scored more than 1 *SD* below the mean of the (a) published norms for the Peabody Picture Vocabulary Test–Revised (PPVT-R; Dunn & Dunn, 1981) and/or (b) local norms (Johnson, Taback, Escobar, Wilson, & Beitchman, 1999) for the Spoken Language Quotient of the Test of Adolescent/Adult Language–3 (TOAL-3; Hammill, Brown, Larsen, & Weiderholt, 1994). Both speaking and listening skills were assessed in the four subtests that constitute the Spoken Language Quotient of the TOAL-3. At age 5, language impairment had been identified using similar criteria and age-appropriate measures (Beitchman et al., 1986).

Procedure

Most OLS participants were tested individually in a face-to-face situation. One individual selected for the current studies participated via telephone because he lived in a distant area.

Single-Word Measures. Three single-word tasks tapped naming and semantic retrieval proficiency. First, the standard score from the Brief Test of the Test of Adolescent/Adult Word Finding (TAWF; German, 1990) reflected the accuracy of naming across 40 total items in several tasks, including confrontation naming (nouns and verbs), category naming, and naming in response to a description. The second single-word measure, total unique words from the Controlled Oral Word Association Task (FAS; Benton & Hamsher, 1978), reflected both the speed and accuracy of word retrieval. In this task, participants produced as many words as possible beginning with a given letter (F, A, or S) within 1 min. The number of unique words produced was totaled. The final single-word measure, the Rapid Automatized Naming task (RAN; Denckla & Rudel, 1976) tapped speed of retrieval. As rapidly as possible, participants named a list of 50 items, comprised of five different digits, presented in a random sequence. RAN time (in seconds) was recorded using a stopwatch, with shorter times reflecting faster word retrieval. A psychometrist administered the RAN and FAS tests; a certified speech-language pathologist administered the TAWF.

Connected Speech Samples. The speech-language pathologist who administered the TAWF also elicited conversational samples using a standard series of interview questions (see Appendix). The speech-language pathologist provided occasional comments to promote a conversational feel to the exchange, rather than just eliciting a monologue from the participant. The conversations were audiotaped and transcribed into the Systematic Analysis of Language Transcripts program (SALT; Miller & Chapman, 1996) by a group of trained university students. The investigators then reviewed the transcripts, segmented them into T-units (Hunt, 1965), and coded conversational disruptions. For reliability purposes, all minimal responses, such as *yes*, *no*, *ok*, and *mhm*, were counted as single-word T-units.

Conversational Disruptions. Conversational disruption types for the present research were identified from those commonly cited in the literature as indications of word-finding difficulty. We initially coded five types of conversational disruptions in the SALT transcripts: empty words, metalinguistic comments, mazes, utterance fillers, and substitutions. Empty words (EMPTY) were words with unspecified referents, such as *thing*. Metalinguistic comments (META) were overt statements of word-finding difficulty (e.g., *What's the word I want?*). Mazes (MAZES) were repetitions or reformulations of words, partial words, or phrases, and filled pauses (e.g., *um, er*). The transcripts were also searched for eight different utterance fillers (FILLERS) including *I don't know*, *you know*, *I mean*, *like*, *well*, *I guess*, *and stuff*, and *whatever*. Substitutions were defined as incorrect words resembling target words in phonetic, semantic, or functional characteristics. Unfortunately, the interview format provided limited shared

context for the interviewer and participant, making it difficult to identify substitutions reliably. Therefore, substitutions were omitted from further analysis. The total numbers of disruptions in each category (EMPTY, META, MAZES, and FILLERS) were tallied for each participant. To control for differences among participants in total words produced, the frequency of each conversational disruption type was then calculated per 100 unmazed words (Dollaghan & Campbell, 1992).

Study 1

The first study evaluated the relationship between single-word task scores and the frequency of conversational disruptions in a young adult population. We expected that individuals who showed fast/accurate responses on single-word tasks should demonstrate few conversational disruptions, whereas those who demonstrated slow/inaccurate responses on single-word tasks should show more frequent disruptions.

Method

Participants

Forty participants (29 males) were selected randomly from the pool of 237 eligible participants in the OLS. The high proportion of males in the Study 1 sample reflects the composition of the original OLS sample (65% male; Beitchman et al., 1986). At age 5, more boys than girls were identified with communication disorders (speech disorders, language disorders, or both). Sex was then one of the criteria used to match the participants with and without disorders who were followed longitudinally in the OLS.

The Study 1 sample also reflects the initial OLS in another way. Specifically, it contains individuals with and without language disorders. Nine participants were judged to have language impairments at age 19. The inclusion of participants with and without language disorders enabled us to represent the full range of possible word-finding skills. The top portion of Table 1 shows descriptive statistics for participants' ages, language scores, and cognitive abilities. Considerable variability in language and cognitive skills is reflected in the large ranges.

Reliability

One investigator transcribed five randomly selected conversational samples. Later, a second investigator listened to the conversational samples, reviewed the transcripts, and recorded any disagreements. The formula used to calculate percentage agreement was: % Agreement = *N* agreements / (*N* agreements + *N* disagreements). Percentages of agreement for word-by-word transcription, T-unit segmentation, and maze coding were 99%, 91%, and 83%, respectively.

Results and Discussion

First, we report descriptive statistics regarding single-word and connected speech measures, particularly the frequencies of conversational disruptions. Second, we examine correlations among the various single-word measures and among the different conversational disruption types. Finally, we report regression analyses predicting conversational disruptions from performance on single-word measures.

TABLE 1. Study 1: Participant (*N* = 40) attributes and descriptive statistics.

Variables	*M*	*SD*	Range
Participant attributes			
Age (years;months)	18;10	0;5	18;5–20;0
TOAL-3 SLQ	97.33	17.80	52–128
Performance IQ (WAIS-R)[a]	104.18	19.24	68–143
PPVT-R standard score	99.60	21.09	40–135
Single-word measures			
TAWF standard score	97.08	19.20	52–139
FAS total unique words	39.90	11.96	19–70
RAN time (s)	17.80	3.79	12.79–26.72
Conversational characteristics			
Total T-units	144.33	53.86	44–322
Total main body words	1196.05	571.39	129–2646
Time (min)	11.61	2.61	4.95–18.03
Mean length of T-unit in words	8.07	2.26	2.82–14.13
Conversational disruptions[b]			
EMPTY/META	0.15	0.17	0–0.78
MAZES	5.71	2.09	2.20–12.40
FILLERS	4.38	1.71	0.90–8.56
COMPZ	0.00	0.71	(–1.37)–(2.95)

[a] Wechsler Adult Intelligence Scale–Revised (WAIS-R; Wechsler, 1981). [b] Frequency per 100 unmazed words.

Descriptive Statistics

Table 1 shows the means, standard deviations, and ranges for single-word measures, conversational characteristics, and frequencies of conversational disruptions. Note that there was a substantial range of scores on each measure.

The average length of the conversations was 11.61 min (*SD* = 2.61). Table 1 also gives summary information on total T-units, mean length of T-units, and total words.

To provide a common metric for comparison, the frequencies for each disruption type were first counted and then expressed as a function of 100 unmazed words. Participants used an average of 0.13 empty words, 0.02 metalinguistic comments, 5.71 mazes, and 4.38 fillers per 100 unmazed words. Because of their low frequencies, the EMPTY and META categories were combined in all remaining analyses (EMPTY/META).

A composite measure (COMPZ) that gave equal weighting to MAZES, FILLERS, and EMPTY/META was calculated. For each participant, the mean frequencies of MAZES, FILLERS, and EMPTY/META were converted to separate *z* scores. The average of the three *z* scores was

recorded as COMPZ ($M = 0.00$, $SD = 0.71$). A composite variable comprising multiple measures of a single construct, such as COMPZ, may be a more reliable, stable, and unbiased estimator of the construct than any of the single measures (Rushton, Brainerd, & Pressley, 1983). Accordingly, a composite may demonstrate stronger correlations with other variables of interest, in this case, the single-word measures.

Correlations Among Single-Word and Receptive Vocabulary Measures

Correlations were computed to assess the relationships among the single-word measures. As shown in Table 2, FAS scores showed a modest but significant positive correlation with TAWF scores ($r = .33$, $p < .05$) and a modest negative correlation with RAN times ($r = –.37$, $p < .05$). That is, higher FAS scores were associated with higher TAWF scores and faster RAN times. RAN times and TAWF scores were not significantly correlated. This pattern of modest correlations among single-word measures suggests that they are measuring relatively distinct abilities rather than a common skill.

Correlations between receptive vocabulary scores on the PPVT-R and the single-word measures were also calculated (see Table 2). Receptive vocabulary scores were included as a possible reflection of the hypothesis that inadequate storage of lexical items may underlie word-finding difficulties (Kail & Leonard, 1986; Nippold, 1992). The PPVT-R standard scores showed strong positive correlations with TAWF standard scores ($r = .71$, $p < .01$) and FAS scores ($r = .61$, $p < .01$), and a small negative correlation with RAN times ($r = –.32$, $p < .05$). These significant correlations indicate that the single-word accuracy measures, in particular, tapped skills that were not independent of those measured by the receptive vocabulary test.

Correlations Among Conversational Disruption Types

Correlations among the conversational disruption types are shown in Table 3. Only a moderate relationship was detected between MAZES and EMPTY/META ($r = .43$, $p < .01$), suggesting that the various conversational disruption types are relatively independent of each other.

TABLE 2. Correlations among single-word measures and composite *z* score (COMPZ) for conversational disruptions.

Measure	1	2	3	4	5
1. TAWF standard score	—	.33*	–.24	.71**	–.17
2. FAS total unique words		—	–.37*	.61**	–.11
3. RAN time			—	–.32*	.34*
4. PPVT-R standard score				—	–.13
5. COMPZ					—

*$p < .05$. **$p < .01$.

Prediction of Conversational Disruptions From Single-Word Measures

A stepwise multiple regression analysis evaluated whether the single-word and receptive vocabulary measures predicted the frequency of conversational disruptions, as indexed by the composite measure, COMPZ. RAN time was the only significant predictor of COMPZ, $F(1, 37) = 4.73$, $p < .05$, accounting for a small amount of variance ($R^2 = .11$). To further investigate the source of this significant finding, separate regression analyses were conducted to predict the frequencies of individual conversational disruption types. RAN time significantly predicted MAZES, $F(1, 37) = 5.12$, $p < .05$, and EMPTY/META, $F(1, 37) = 4.32$, $p < .05$, accounting for small amounts of variance for each conversational disruption type ($R^2 = .12$ and .11, respectively). No other predictors were significantly associated with individual conversational disruption types.

Impaired Performances on Single-Word Measures

A supplementary analysis determined how many individuals in the Study 1 sample showed single-word scores that might be reflective of impairment on these tasks. Local norms for the RAN and FAS tasks were developed from the entire OLS sample, using a statistical weighting procedure (Johnson, Taback, et al., 1999). Published norms were available for the TAWF. Using cutoff scores of 1 *SD* below the mean, 14 of 40 participants demonstrated one or more single-word scores suggestive of impaired performance. Eight of those 14 had language disorders at the time of testing.

Contrary to expectations from the word-finding literature, individuals with poor single-word scores did not clearly demonstrate more frequent conversational disruptions than those with good single-word scores. The primary associations observed were modest ones between slow RAN times and increased frequencies of MAZES and EMPTY/META in conversation. A possible interpretation is that these disruption types may, in part, reflect delays in word retrieval. The weak nature of the relationships, however, suggests that other factors probably also underlie the production of these disruptions. Moreover, receptive vocabulary scores also did not predict the frequencies of conversational disruptions, as might be expected from a storage elaboration account of word-finding difficulties (Kail & Leonard, 1986).

TABLE 3. Correlations among frequencies of conversational disruptions per 100 words.

Conversational Disruptions	1	2	3
1. EMPTY/META	—	.43**	.30
2. MAZES		—	.04
3. FILLERS			—

**$p < .01$.

Other expectations from the word-finding literature also received limited support. For example, the various single-word measures used here did not relate strongly to each other as would be expected if they tapped a common ability. Similarly, the various types of conversational disruptions showed weak interrelationships, inconsistent with a unitary underlying construct. An alternate possibility is that different speakers use different conversational disruption types to cope with underlying word-finding difficulties, in which case correlations among measures would not be expected. One could also argue that perhaps the disruption measures chosen here were not the best choices to assess word-finding behaviors. They were, however, representative of the types of measures recommended in the literature (German, 1987; German & Simon, 1991; Heller & Dobbs, 1993; Jordan et al., 1997; Snyder & Godley, 1992).

We assumed in Study 1 that word-finding proficiency is normally distributed within the population (Heller & Dobbs, 1993). The assumption represents a reasonable extension from the literature according to the following argument. Word-finding problems are claimed to be a common characteristic of individuals with developmental language impairments (e.g., Faust et al., 1997; Nippold, 1992; Wiig & Semel, 1984). Individuals with developmental language impairments may represent the lower end of the normal distribution with respect to various language skills (Leonard, 1998), including perhaps word-finding ability. Our sample contains participants both with and without developmental language disorders. These participants showed wide ranges of performance on the relevant measures, ranges sufficient to reveal robust correlations between single-word performances and conversational fluency if they existed. Moreover, a substantial number of participants in our sample showed single-word scores that could be considered indicative of impairment on those tasks.

It is possible, however, that our Study 1 assumption of a normal distribution of word-finding abilities was not correct. Thus, in Study 2, we took a different approach to selection of participants in an attempt to further evaluate possible relationships between single-word scores and conversational fluency.

Study 2

Study 2 compared the single-word scores of two groups that differed in subjective ratings of conversational fluency. It was expected that if a relationship exists between single-word task performance and conversational fluency, the group with poor fluency ratings would demonstrate lower single-word scores than the group with adequate fluency ratings.

Method

Participants

Participants from Study 1 were returned to the OLS pool of 237 eligible individuals before the selection of participants for Study 2. Ratings of conversational fluency were used to select 40 participants for Study 2. After conducting the OLS conversational interview, the speech-language pathologist provided a rating of perceived conversational fluency for each participant. One of four possible rating categories was assigned to complete the statement, *The rhythm of speech is*: (a) normally fluent and does not attract listener attention, (b) characterized by frequent pauses and linguistic revisions but few "stutter-like" disfluencies (e.g., blocks, part- or whole-word repetitions, struggle behaviors), (c) interrupted by occasional "stutter-like" disfluencies, or (d) interrupted by frequent "stutter-like" disfluencies. As mentioned earlier, participants who stuttered (categories c and d) were excluded from the present studies. Participants ranked in the (a) category were considered to have adequate conversational fluency; those in the (b) category were considered to have poor conversational fluency.

For Study 2, the total of 40 participants included 20 speakers (of whom 18 were males) judged to have poor conversational fluency (Group P) and 20 others judged to have adequate conversational fluency (Group A). The two groups were matched on sex and on PPVT-R standard scores at age 19. The top panel of Table 4 shows descriptive statistics for participants' ages, language skills, and cognitive abilities. The two groups did not differ significantly in performance IQs or TOAL-3 Spoken Language Quotients. Sixteen participants had language impairments (8 in each fluency group) at age 19. Thirteen participants in Study 2 were also involved in Study 1.

Reliability

Two investigators reviewed five randomly selected transcripts (different from the transcripts reviewed in Study 1) following the procedures outlined in Study 1. Agreement percentages were 98% for word-by-word transcription, 89% for T-unit segmentation, and 87% for maze coding.

Results and Discussion

We first assessed group differences in single-word performance, followed by differences in conversational disruptions.

Group Differences in Single-Word Task Performance

Table 4 shows the single-word scores for the two groups. Participants with poor and adequate ratings of conversational fluency did not differ significantly on any of the single-word measures, suggesting that such tasks do not tap the same abilities considered when making clinical judgments of conversational fluency. Twenty-five of the 40 participants showed performances in the impaired range (<1 *SD* below mean) on one or more of the single-word tasks, 13 from the group with poor fluency ratings and 12 from the group with adequate ratings. As in Study 1, these results call into question the presumed relationships between proficiency on single-word tasks and conversational fluency.

TABLE 4. Study 2: Participant attributes and descriptive statistics as a function of conversational fluency group (*N* = 20 per group).

	Conversational Fluency Group					
	Poor (P)			Adequate (A)		
Variables	*M*	*SD*	Range	*M*	*SD*	Range
Participant attributes						
Age (years;months)	18;9	0;3	18;5–19;5	19;0	0;5	18;6–19;9
TOAL-3 SLQ	91.60	12.81	72–113	89.90	12.39	67–123
Performance IQ (WAIS-R)[a]	101.25	15.93	76–131	99.95	15.58	76–132
PPVT-R standard score	92.60	13.50	70–128	92.60	14.05	70–130
Single-word measures						
TAWF standard score	92.90	18.30	71–139	93.80	13.86	74–113
FAS total unique words	32.05	8.77	19–48	36.95	12.23	15–59
RAN time (s)	18.67	4.15	12.79–26.30	19.39	4.17	14.05–29.83
Conversational characteristics						
Total T-units	160.30	64.69	67–301	137.25	39.43	65–209
Total main body words	1292.75	689.60	312–2929	1201.80	443.22	552–1970
Time (min)	12.79	4.44	8.42–26.58	11.43	2.74	7.68–17.08
Mean length of T-unit (words)	7.94	2.22	4.11–11.54	8.85	2.31	5.78–14.13
Conversational disruptions[b]						
EMPTY/META	0.23	0.24	0–0.75	0.12	0.12	0–0.48
MAZES	6.57	2.31	2.73–11.22	6.01	2.77	3.38–15.28
FILLERS	5.65	1.94	2.05–10.27	4.84	2.15	2.09–9.24
COMPZ*	0.19	0.66	(–0.82)–(1.62)	–0.19	0.62	(–1.08)–(1.40)

[a] Wechsler Adult Intelligence Scale–Revised (WAIS-R; Wechsler, 1981).
[b] Frequency per 100 unmazed words.
*$p < .05$ one-tailed.

Group Differences in Conversational Disruptions

We then tested the validity of our participant groupings by assessing whether the groups formed on the basis of subjective ratings of conversational fluency actually differed in their objective frequencies of conversational disruptions, as measured by the composite measure COMPZ. Participants judged to have poor conversational fluency had higher COMPZ scores, reflective of more frequent conversational disruptions, than those judged to have adequate conversational fluency, $t(19) = 2.06$, $p < .05$ one-tailed. The effect size for the group difference in the COMPZ measure of objective disruptions ($d = .58$) was moderate (Cohen, 1988), despite the restricted range of the subjective ratings (0, 1) that had been used to determine the group assignments.

General Discussion

Two studies with different methodologies provided converging evidence for, at best, weak relationships between single-word task performances and frequencies of conversational disruptions. This conclusion is consistent with evidence from the small number of studies that have assessed this relationship (Heller & Dobbs, 1993; Jordan et al., 1997). The conclusion, however, is not consistent with the traditional clinical argument that individuals who perform poorly on single-word tasks will show frequent conversational disruptions because both are reflections of underlying word-finding deficits (German, 1987; German & Simon, 1991). Our results show that some who perform poorly on single-word tasks will show few conversational disruptions and some who perform well on single-word tasks will show frequent conversational disruptions. A few individuals will perform poorly on single-word tasks *and* show frequent conversational disruptions. What remains to be determined is whether these cases reflect chance co-occurrences of both difficulties or the common influence of a single underlying construct such as word finding.

Our studies suggest that further work is required to define and validate the construct of word-finding difficulties, particularly in the developmental population. The construct was likely borrowed from the literature on acquired language disorders, where there is compelling evidence that brain damage leads to obvious changes in naming and retrieval proficiency in both single-word and conversational situations (Goodglass, Kaplan, Weintraub, & Ackerman, 1976; Marshall, 1976). As with other constructs borrowed from the acquired literature, such as developmental apraxia of speech, it is challenging but

crucial to validate the applicability of the word-finding construct to those with developmental language problems where dramatic onset and identifiable etiology are not evident. It needs to be shown empirically that the construct designates a group of individuals with consistent, unique patterns of behavior that warrant specific interventions.

Our studies do not rule out the possibility that certain individuals demonstrate a pattern of related deficits in both single-word and conversational tasks, as expected from traditional ideas about word-finding problems. Perhaps these individuals are relatively rare, even among young adults identified with language impairments. Our studies may have been based on samples of participants too small or too diverse (with and without language impairments) to include a sufficient number of individuals with this particular pattern of deficits.

Part of the current uncertainty is that there is no accepted common standard for the diagnosis of word-finding deficits in individuals with developmental language impairments (Dockrell, Messer, George, & Wilson, 1998). Four types of measures are commonly employed: performance on single-word naming/retrieval tasks, performance on receptive vocabulary measures (presumably to verify intact comprehension), subjective assessments of conversational fluency, and objective assessments of conversational fluency. Let us consider, in turn, some of the difficulties inherent in each type of measure.

Single-word tasks, such as naming and lexical retrieval, involve complex cognitive behaviors thought to entail multiple steps (e.g., Johnson, Paivio, & Clark, 1996; Levelt, Roelofs, & Meyer, 1999) and diverse processes: semantic, syntactic, phonological, and motor. Thus, inaccurate or slow performances may result from multiple processing inefficiencies (Lahey & Edwards, 1996). Single-word tasks are also influenced by numerous experiential and contextual factors (Johnson et al., 1996). Moreover, careful cognitive analysis and the correlations reported here suggest that the various single-word tasks often recommended for assessment of word-finding deficits do not necessarily tap common abilities.

Lexical storage difficulties may also affect performance on single-word tasks (Faust et al., 1997; Kail & Leonard, 1986; McGregor, Friedman, Reilly, & Newman, 2002). As shown here, receptive vocabulary scores are strongly related to scores on certain types of single-word tasks, making it difficult to disentangle storage versus retrieval processes, both empirically and conceptually. Typical receptive vocabulary tasks may also be relatively insensitive to subtle differences in storage and elaboration of word knowledge (McGregor & Leonard, 1995), further complicating the interpretation of possible word-finding difficulties and their origins.

A possible limitation of the current studies concerns the single-word tasks that were used. The choices were restricted to measures that had already been collected as part of the OLS. Although the measures were ones that are commonly mentioned in the word-finding literature, it is possible that other single-word tasks would yield a more appropriate or sensitive assessment of underlying word-finding difficulties.

Our understanding of conversational disruptions is even more limited than our understanding of single-word difficulties. Reliable, objective counts of disruptions to conversational fluency are time consuming and difficult to obtain (Dollaghan & Campbell, 1992). Consequently, clinicians have turned to subjective judgments of conversational fluency, such as checklists of conversational disruptions, to help identify those who may be experiencing word-finding difficulties (German, 1983). A common diagnostic battery for word-finding disorders includes various naming or semantic retrieval tasks and a subjective checklist of conversational disruptions. We showed in Study 2 that the speech-language pathologist's subjective ratings of fluency were predictive of differences in objective measures of conversational fluency. German and colleagues (German, 1987; German & Glasnapp, 1990; German & Simon, 1991) have also supplied evidence that can be interpreted as supportive of a relationship between subjective and objective measures of conversational fluency, although further exploration of the determinants of this relationship would be desirable.

An intriguing possibility is that clinician judgments of word-finding difficulties may be less affected by the sheer number of conversational disruptions than by the impact those disruptions have on listener comprehension (Dollaghan & Campbell, 1992). Future research to disentangle quantitative and qualitative aspects of conversational disruptions may advance our thinking and practice concerning these issues.

It is conceivable that the conversational interview questions we used did not sufficiently tax participants' word-finding skills because low-frequency vocabulary items or complex syntactic structures were not consistently required. However, a supplementary analysis of responses to the opinion/challenge question, the most difficult portion of the interview, showed results similar to those reported here. Nonetheless, tasks constraining the speaker to use more specific vocabulary and more complex syntax, such as narrative retelling tasks, may provide a more rigorous test of the relationship between single-word and connected speech measures.

Further work is also needed to understand the factors responsible for conversational disruptions, such as mazes, empty words, metalinguistic comments, and utterance fillers. Among the candidate factors are pragmatic demands (Heller & Dobbs, 1993), emotional state (Snyder & Godley, 1992), syntactic complexity (MacLachlan & Chapman, 1988), and topic knowledge (Perry & Lewis, 1999). We need to understand how these complex factors interact to produce disruptions in speakers both with and without various communication disorders (e.g., word-finding deficits, stuttering). Concerted efforts will be required to address these issues, which are critical to the science and practice of speech-language pathology.

In the interim, findings from these preliminary investigations suggest that clinicians should exercise caution in the identification and treatment of possible word-finding deficits associated with developmental language disorders. Single-word measures alone cannot be used to infer deficits in conversational fluency, nor can conversational

fluency measures alone (either subjective or objective) be used to infer deficits on single-word tasks. Likewise, gains made in intervention at the level of single-word tasks may not be observed at a more functional level such as conversation and vice versa. Thus, clinicians should maintain a healthy skepticism about the existence of word-finding deficits as discrete problems readily distinguished from other developmental language difficulties (Dockrell et al., 1998; McGregor & Leonard, 1995).

Acknowledgments

This article is based on a Master's research project completed by the first author under the supervision of the third author. Portions of this work were presented at the 1998 Annual Convention of the American Speech-Language-Hearing Association in San Antonio, TX, and the 2000 meeting of the Tri-Joint Congress in Toronto, Ontario, Canada. This research was supported by Grant 6606-5639-102 to Joseph H. Beitchman and colleagues from the National Health Research Development Program, Health and Welfare Canada, and by a grant to Carla J. Johnson from the Natural Sciences and Engineering Research Council of Canada.

References

Beitchman, J. H., Nair, R., Clegg, M., & Patel, P. G. (1986). Prevalence of speech and language disorders in 5-year-old kindergarten children in the Ottawa–Carleton region. *Journal of Speech and Hearing Disorders, 51*, 98–110.

Benton, A. L., & Hamsher, K. (1978). *Multilingual Aphasia Examination*. Iowa City: University of Iowa Hospitals.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Denckla, M. B., & Rudel, R. G. (1976). Rapid "Automatized" Naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14,* 471–479.

Dockrell, J. E., Messer, D., George, R., & Wilson, G. (1998). Children with word-finding difficulties—prevalence, presentation, and naming problems. *International Journal of Language and Communication Disorders, 33,* 445–454.

Dollaghan, C. S., & Campbell, T. F. (1992). A procedure for classifying disruptions in spontaneous language samples. *Topics in Language Disorders, 12*(2), 56–68.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test–Revised*. Circle Pines, MN: American Guidance Service.

Faust, M., Dimitrovsky, L., & Davidi, S. (1997). Naming difficulties in language-disabled children: Preliminary findings with the application of the tip-of-the-tongue paradigm. *Journal of Speech, Language, and Hearing Research, 40,* 1026–1036.

German, D. J. (1983). I know it but I can't think of it: Word retrieval difficulties. *Academic Therapy, 18,* 539–545.

German, D. J. (1987). Spontaneous language profiles of children with word-finding problems. *Language, Speech, and Hearing Services in Schools, 18*, 217–230.

German, D. J. (1990). *Test of Adolescent/Adult Word Finding*. Chicago: Riverside.

German, D. J. (1992). Word-finding intervention for children and adolescents. *Topics in Language Disorders, 13,* 33–50.

German, D. J., & Glasnapp, D. R. (1990). The Test of Word Finding in Discourse: Diagnostic utility evidence. *Educational and Psychological Measurement, 50,* 383–392.

German, D. J., & Simon, E. (1991). Analysis of children's word-finding skills in discourse. *Journal of Speech and Hearing Research, 34,* 309–316.

Goodglass, H., & Kaplan, E. (1972). *The assessment of aphasia and related disorders*. Philadelphia: Lea & Febiger.

Goodglass, H., Kaplan, E., Weintraub, S., & Ackerman, N. (1976). The "tip-of-the-tongue" phenomenon in aphasia. *Cortex, 12,* 145–153.

Goodglass, H., Wingfield, A., & Hyde, M. R. (1998). The Boston corpus of aphasic naming errors. *Brain and Language, 64*, 1–27.

Hall, P. K., & Jordan, L. S. (1987). An assessment of a controlled association task to identify word-finding problems in children. *Language, Speech, and Hearing Services in Schools, 18*, 99–111.

Hammill, D., Brown, V., Larsen, S., & Weiderholt, J. (1994). *Test of Adolescent/Adult Language–3*. Austin TX: Pro-Ed.

Heller, R. B., & Dobbs, A. R. (1993). Age differences in word finding in discourse and nondiscourse situations. *Psychology and Aging, 8*, 443–450.

Hunt, K. (1965). *Grammatical structure written at three grade levels* (Research Report No. 3). Urbana, IL: National Council of Teachers of English.

Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., et al. (1999). Fourteen year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language, and Hearing Research, 42,* 744–760.

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin, 120,* 113–139.

Johnson, C. J., Taback, N., Escobar, M., Wilson, B., & Beitchman, J. H. (1999). Local norming of the Test of Adolescent/Adult Language–3 (TOAL-3) in the Ottawa speech and language study. *Journal of Speech, Language, and Hearing Research, 42,* 761–766.

Jordan, F., Ward, K., & Cremona-Meteyard, S. (1997). Word-finding in the conversational discourse of children with closed head injury. *Aphasiology, 11*, 877–888.

Kail, R., & Leonard, L. B. (1986). Word-finding abilities in language-impaired children. *ASHA Monographs, 25*.

Knopman, D. S., Selnes, O. A., Niccum, N., & Rubens, A. B. (1984). Recovery of naming in aphasia: Relationship to fluency, comprehension and CT findings. *Neurology, 34*, 1461–1470.

Lahey, M., & Edwards, J. (1996). Why do children with specific language impairment name pictures more slowly than their peers? *Journal of Speech and Hearing Research, 39,* 1081–1098.

Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, MA: Bradford.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–75.

MacLachlan, B. G., & Chapman, R. S. (1988). Communication breakdowns in normal and language learning-disabled children's conversation and narration. *Journal of Speech and Hearing Disorders, 53,* 2–7.

Marshall, R. C. (1976). Word retrieval behavior of aphasic adults. *Journal of Speech and Hearing Disorders, 41,* 444–451.

McGregor, K. K., Friedman, R. M., Reilly, R. M., & Newman, R. M. (2002). Semantic representation and naming in young children. *Journal of Speech, Language, and Hearing Research, 45*, 332–346.

McGregor, K. K., & Leonard, L. B. (1995). Intervention for word-finding deficits in children. In M. E. Fey, J. Windsor, & S. F. Warren (Eds.), *Language intervention: Preschool*

through the elementary years (pp. 85–105). Baltimore: Paul H. Brookes.
Miller, J., & Chapman, R. S. (1996). *Systematic Analysis of Language Transcripts*. Madison: Waisman Center, University of Wisconsin–Madison.
Nippold, M. A. (1992). The nature of normal and disordered word finding in children and adolescents. *Topics in Language Disorders, 13*(1), 1–14.
Perry, M., & Lewis, J. L. (1999). Verbal imprecision as an index of knowledge in transition. *Developmental Psychology, 35,* 749–759.
Rapin, I., & Wilson, B. (1978). Children with developmental language disability: Neurological aspects and assessment. In M. Wyke (Ed.), *Developmental dysphasia* (pp. 13–41). New York: Academic Press.
Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin, 94,* 18–38.
Scott, C., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research, 43,* 324–339.
Snyder, L. S., & Godley, D. (1992). Assessment of word-finding disorders in children and adolescents. *Topics in Language Disorders, 13*(1), 15–32.
Wechsler, D. (1981). *Wechsler Adult Intelligence Scale–Revised.* New York: The Psychological Corporation.
Wiig, E. H., & Semel, E. M. (1975). Productive language abilities in learning disabled adolescents. *Journal of Learning Disabilities, 8*, 578–586.
Wiig, E. H., & Semel, E. M. (1984). *Language assessment and intervention for the learning disabled*. Columbus, OH: Merrill.

Received November 28, 2001
Accepted February 14, 2003
DOI: 10.1044/1058-0360(2003/081)

Contact author: Carla J. Johnson, PhD, Graduate Department of Speech-Language Pathology, Faculty of Medicine, University of Toronto, Rehabilitation Sciences Building, 500 University Avenue, Toronto, Ontario M5G 1V7, Canada. E-mail: carla.johnson@utoronto.ca

Appendix

Conversational Interview Questions

1. Perhaps you could start by telling me a bit about yourself. You could tell me about your family, what you do everyday, or any special interests you have.
2. If I were to ask your friends about you, how do you think they would describe you to me? (If necessary, probe further by asking "What unique qualities or special interests and abilities would they tell me about?").
3. Is there anything that you have done in your life that you would change or do differently if you could do it over?
4. I feel that it's a very different world than it was 20 years ago. I'd like to know your personal views about how best to get along in today's world. What is the key to being successful these days?
5. I'd also like to get your opinion about a controversial issue that you feel strongly about. I have some suggestions for you, or you can use any topic you choose. My suggestions are environmental issues, abortion, euthanasia, cuts to welfare, or the Young Offender's Act. But I want you to choose something you feel strongly about and would like to comment on. [After participant gave opinion, SLP challenged the participant's opinion by expressing an opposite view (e.g., Some people would say … How would you respond to that?)]