Atypical pragmatic language is often present in individuals with autism spectrum disorders (ASD), along with delays or deficits in structural language. This study investigated the use of the “fillers” \textit{uh} and \textit{um} by children ages 4–8 during the autism diagnostic observation schedule. Fillers reflect speakers’ difficulties with planning and delivering speech, but they also serve communicative purposes, such as negotiating control of the floor or conveying uncertainty. We hypothesized that children with ASD would use different patterns of fillers compared to peers with typical development or with specific language impairment (SLI), reflecting differences in social ability and communicative intent. Regression analyses revealed that children in the ASD group were much less likely to use \textit{um} than children in the other two groups. Filler use is an easy-to-quantify feature of behavior that, in concert with other observations, may help to distinguish ASD from SLI.

**Introduction**

Language abilities in children with autism spectrum disorders (ASD) are highly variable [Tager-Flusberg and Joseph, 2003; Whitehouse et al., 2008] although delays and deficits are relatively common [Leyfer et al., 2008; Loucas et al., 2008]. Recent studies suggest that a majority of verbally fluent children with ASD have impairments in \textit{structural language}, which includes phonology, vocabulary, and grammar. In contrast, \textit{pragmatic language}—the socially-oriented elements of language use—is thought to be universally impaired in ASD [Kim et al., 2014; Klin et al., 2005; Landa, 2000; Lord and Paul, 1997; Tager-Flusberg et al., 2005; Volden et al., 2009]. Yet, there is little consensus on how pragmatic language abilities should be defined or quantified [Russell and Grizzle, 2008; Volden and Phillips, 2010]. More generally, speech communication is critical for everyday functioning, so there is great potential value for interventions that might increase the capacity of an individual with ASD to understand and be understood [Klin et al., 2007].

In this study, we investigated one quantifiable feature of pragmatic language: the use of \textit{uh} and \textit{um}. These “fillers” (or “filled pauses”) are subtle yet ubiquitous features of spontaneous speech, accounting for approximately one percent of word tokens produced by typical adults [Acton 2011; see also Fox Tree, 1995, p. 709]. Like other types of disfluency (including pauses, false starts, repetitions, and repairs), fillers are thought to reflect difficulties with planning and delivering speech [Clark, 1994; Levelt, 1989, p. 484]. For instance, fillers are particularly common immediately before pauses in speech [Clark and Fox Tree, 2002]. But there is extensive evidence that fillers also act as interpersonal displays directed at the listener [Sacks, Schegloff, & Jefferson, 1974]. On hearing an \textit{uh} or \textit{um}, listeners may infer that the speaker is experiencing difficulty with word retrieval or speech planning [Maclay and Osgood, 1959; Stenström, 1994, p. 76f.], and will often provide verbal assistance to the speaker [Clark and Wilkes-Gibbs, 1986; Jefferson, 1974]. Listeners may also use fillers to uncover linguistic structure during speech perception. For instance, fillers may be used as cues to major syntactic boundaries [Bailey and Ferreira, 2003; Martin and Strange, 1968; Swerts, 1998] or to new information being introduced into the discourse [Arnold et al., 1998].
Listeners also may use fillers as cues to the speaker’s mental state; for example, high rates of filler use often causes listeners to infer that the speaker lacks knowledge about the topic under discussion [Brennan and Clark, 1995; Smith and Clark, 1993]. Thus, regardless of the speaker’s intent, fillers can influence how a listener perceives and responds to a speaker during conversation.

In summary, filler use appears to be an important component of conversational reciprocity, which is thought to be impaired in ASD [e.g., Tager-Flusberg et al., 2005]; for example, compared to typically developing (TD) children, children with ASD have difficulty initiating conversation [Tager-Flusberg, 1996], responding appropriately to the initiations of others [Adams et al., 2002; Capps et al., 1998; Stone and Caro-Martinez, 1990], taking turns [Bottig and Conti-Ramsden, 2003; Ramberg et al., 1996], and staying on topic [Capps et al., 1998; Lam and Yeung, 2012; Losh and Capps, 2003; Loveland et al., 1990; Paul et al., 2009]. We thus hypothesized that difficulties with conversational reciprocity would also be reflected in atypical filler use in children with ASD.

Other social-cognitive impairments in ASD may also contribute to atypical filler use, for instance executive functioning difficulties, which are common in individuals with ASD [Kenworthy, Yerys, Anthony, & Wallace, 2008]. One executive function (inhibition) has been linked to disfluency in typical adults [Engelhardt et al., 2013] and other studies of typical adults have found that cognitive load increases disfluency rates [e.g., Bortfeld et al., 2001; Christenfeld, 1995; Engelhardt et al., 2010; Schachter et al., 1991]; presumably, cognitive load attenuates speakers’ ability to monitor their speech planning for anticipated delays [Bock, 1982; Levelt, 1983]. Adams et al. [2002] suggests that “social-emotional conversation” may be particularly loading for children with ASD. In that study, children with Asperger syndrome produced more “pragmatically problematic” responses (roughly, those judged to be pragmatically inappropriate in context) than peers with conduct disorder, but only during social-emotional conversations. Taken together, these studies suggest that filler use in ASD may vary as a function of the social demands of the topic under discussion.

A number of studies have attempted to confirm a clinical impression that high-functioning children with ASD “may lack in fluency” [Klin et al., 2005, p. 99] compared to TD children. Thurber and Tager-Flusberg [1993] examined disfluency in ten children and adolescents with autism during narration of a wordless picture book. Participants with autism produced fewer within-phrase pauses than participants with typical development or mild mental retardation, but no group differences in repetitions or false starts were found. In another study, Lake et al. [2011] elicited conversational speech from 13 adults with ASD. Compared to age-matched controls, ASD participants produced more disfluent repetitions and more pauses, but fewer revisions and fillers. Suh et al. [2014] used a storytelling task to elicit speech from children with high-functioning ASD, children with a past diagnosis of ASD who no longer met criteria for ASD [see Suh et al., 2014 for details], and TD children. They found that both clinical groups produced higher rates of repetitions and revisions than TD children, but filler rates were comparable across groups. These conflicting results are likely due to methodological differences. For example, Suh et al. studied children using a storytelling task to elicit speech, whereas Lake et al. elicited speech by asking their adult participants about their interests and hobbies. In addition, neither study controlled for participants’ language abilities. Indeed, Lake et al. found that one measure of language ability—mean length of utterance—was correlated with disfluency rate, and similar results have been found in studies of typical adults [Bortfeld et al., 2001; Cook, Smith, & Laljlee, 1974; Oviatt, 1995; Shriberg, 1996] and children with specific language impairment (SLI) [Thordardottir & Weismher, 2002]. Thus, group differences may have been due to the lower average language abilities of the ASD group rather than to any specific feature of ASD. In summary, it is still unclear whether children with ASD use fillers differently than TD children or children with language delays not related to ASD.

Although the above studies conflated the fillers *uh* and *um*, Clark and Fox Tree [2002] argue that they have different functions: *uh* serves to signal minor delays, whereas *um* signals major delays. The evidence for this is primarily distributional. In typical adults, *um* is more often followed by a pause than *uh*, and when a pause is present after a filler, it tends to be longer after *um* than after *uh*. Children with ASD also exhibit this pattern, producing more pauses, and longer pauses, after *um* than after *uh* [de Villiers, 2011; Lunsford, Heeman, Black, & van Santen, 2010]. Clark and Fox Tree also report that *um* is the more common of the two fillers at the start of intonational phrases (e.g., *um, we went to the beach*)—where speech planning demands are presumably at their greatest—whereas *uh* is more common elsewhere (*we went to the, uh, beach*). In summary, *uh* and *um* have different usage patterns, and perhaps, different functions in discourse.

**The Present Study**

Heeman et al. [2010], Lunsford et al. [2010], and Lunsford [2012] hypothesized that children with ASD would produce atypical patterns of filler use compared to TD children, and preliminary analyses conducted on a
smaller subgroup of children included in the current sample provided initial support for this hypothesis. However, these prior analyses did not address specific hypotheses regarding the effects of topic on filler use, and did not examine associations between filler use and individual differences in cognitive abilities or ASD symptoms.

In the present study, we re-examined this hypothesis using a larger sample of children with and without ASD, according to best estimate clinical (BEC) diagnoses. For comparison, we included a TD group as well as a group of children with SLI, a neurodevelopmental disorder defined by language delays or deficits in the absence of other developmental or sensory impairments [Tomblin, 2011]. SLI is associated with deficits in structural language whereas ASD involves atypicalities in both structural and pragmatic language [Shulman and Guberman, 2007]. To determine whether there is a specific ASD-related profile for filler use, the SLI clinical group is essential [Bishop, 2001; Ellis Weismer, 2013; Kjelgaard and Tager-Flusberg, 2001]; otherwise, observed group differences may be attributed to difficulties with language that are common in, but not specific to, children with ASD.

**Methods**

**Participants**

One hundred and ten children from the Portland, OR metropolitan area, between 4 and 8 years of age, took part in the study: 50 children with ASD (45 male), 43 TD children (TD; 31 male), and 17 children with SLI (11 male).

**Recruitment and screening.** Participants with ASD were recruited through local healthcare specialists, educational service districts, autism clinics, parent groups, and local nonprofit autism organizations. Participants with SLI were recruited through local speech clinics, speech-language pathologists, and the Oregon Speech and Hearing Association. Advertisements were also placed in local newspapers and on “community tables” at local elementary schools. All participants had full-scale IQ scores of 70 or higher on the Wechsler Preschool and Primary Scale of Intelligence [WPPSI-III; Wechsler, 2002] for children under 7 years of age, or the Wechsler Intelligence Scale for Children [WISC-IV; Wechsler, 2004] for children ages 7 or older. Children were excluded if they had any of the following: (a) known metabolic, neurological, or genetic disorder, (b) gross sensory or motor impairment, (c) brain lesion, (d) orofacial abnormalities (such as cleft palate), or (e) mental retardation. All participants spoke English as their first language, and produced a mean length of utterance in morphemes (MLUM) of at least three. During the initial screening, a certified speech-language pathologist confirmed the absence of speech intelligibility impairments.

**Diagnostic groups.** BEC judgment by experienced clinicians is thought to be the gold standard for ASD diagnosis [e.g., Klin et al., 2000; Spitzer and Siegel, 1990]. In this study, a panel of clinicians, including two clinical psychologists, a speech-language pathologist, and an occupational therapist, all of whom had clinical expertise with ASD, based their judgments on DSM-IV-TR criteria [American Psychiatric Association, 2000] for ASD. Only children who received a consensus BEC diagnosis of ASD were included in this study. The consensus diagnosis was confirmed by above-threshold scores on the Autism Diagnostic Observation Schedule-Generic [ADOS-G; Lord et al., 2000] according to the revised algorithms [Gotham et al., 2007] and the Social Communication Questionnaire [SCQ; Rutter et al., 2003] according to the cutoff score of 12 recommended for research purposes [Lee et al., 2007]. There were a small number of nonresponses on questions on the SCQ (corresponding to 0.3% of the overall data). Before computing SCQ scores, chained equation multiple imputation [Su et al., 2011] was used to fill in nonresponses (corresponding to 0.3% of the overall data).

Language impairment was assessed using the Clinical Evaluation of Language Fundamentals ( CELF), a test which produces a composite summary of expressive and receptive language abilities. For children younger than 6 years of age, the CELF Preschool-2 [Semel, Wiig, & Secord, 2004] was administered; the CELF-4 [Semel et al., 2003] was used for children age 6 or older. Language impairment was determined by a CELF core language score (CLS) more than one standard deviation below the mean. Half of the 50 children with ASD were identified as language impaired according to this criterion. Children assigned to the SLI group also were required to have (a) a documented history of language delays or deficits, and (b) a BEC consensus judgment of language impairment but not ASD, taking into account medical and family history, assessments performed as part of this study or at an earlier time by others, and school records. Children with a BEC diagnosis of SLI were excluded from the study if they reached threshold on both the ADOS-G and the SCQ.

Children who did not meet the criteria for either ASD or SLI were assigned to the TD group, but were excluded from the study if they had any family members diagnosed with either ASD or SLI, a history of psychiatric disturbance (e.g., ADHD) or if the child was above threshold according to the ADOS-G or the SCQ.

**Procedures**

Participants completed a battery of experimental tasks and cognitive, language, and neuropsychological assessments over six sessions of 2–3 hr each. All procedures were approved by the Oregon Health & Science
University Institutional Review Board. Participating families were fully informed about the study procedures and provided written consent.

**Standardized measures.** The ADOS [Lord et al., 2000], a semistructured autism diagnostic observation, was administered to all children in the current study, and was scored according to the revised algorithms [Gotham et al., 2007]. Ten children received ADOS Module 2, and 100 received Module 3. Domain calibrated severity scores (CSS) were calculated as indications of severity of social affect (SA) and restricted and repetitive behavioral (RRB) symptoms (Hus et al., in press). The social affect calibrated severity scores (ADOS SA; range: 1–10) was used as a clinician-reported measure of social communication difficulties. Transcripts of the ADOS were used to derive several other measures (see next section).

Verbal IQ (VIQ), performance IQ (PIQ), and full-scale IQ (FSIQ) were estimated using the Wechsler scales tests, as described above.

Parents completed the behavior rating inventory of executive function [BRIEF; Gioia et al., 2000] for children 6 years of age or older, and the BRIEF-Preschool Version [BRIEF-P; Gioia et al., 2003] for children under 6. Both forms were used to compute the global executive composite (GEC), which was used as a measure of overall executive functioning.

Structural language abilities in children with ASD and SLI were assessed using the CELF CLS, as well as the two CELF subscales, the expressive language index (ELI) and the receptive language index (RLI). TD children were screened for language impairment but did not complete the CELF.

Parents completed the Children’s Communication Checklist [CCC-2; Bishop, 2003], a 70-item questionnaire assessing the child’s communication abilities in natural settings. The general communication composite (GCC) is the sum of subscale scores from the eight CCC-2 domains related to communication (speech, syntax, semantics, coherence, initiation, scripted language, context, and nonverbal communication). The social-interaction deviance index (SIDI) uses these subscales to measure relative strengths in structural vs. pragmatic language; a negative SIDI indicates stronger structural language abilities, and a positive score indicates stronger pragmatic language abilities.

Parents also completed the Social Communication Questionnaire [Rutter et al., 2003], a 40-item assessment of symptomatology associated with ASD. The SCQ communication total score [SCQ-CTS; range: 0–12; Berument et al., 1999], the sum of scores for items in the communication domain, was used as an additional parent-reported measure of communication abilities.

**Transcription.** ADOS sessions were recorded and the child and examiner’s speech was transcribed using Praat software. Transcribers were blind to study hypotheses and to participants’ diagnostic status and intellectual abilities. The transcriptions were generated using a subset of the systematic analysis of language transcripts (SALT) guidelines [Miller and Chapman, 1985]. Transcribers were instructed to mark mazes (i.e., disfluent intervals of speech), including sequences of fillers and false starts, repetitions, and revisions. Each ADOS transcription was segmented into four “activities”: Play (including Make-Believe Play and Joint Interactive Play), Description of a Picture, Telling a Story from a Book, and Conversation. For children who received the ADOS Module 3, the Conversation activity included the Emotions, Social Difficulties and Annoyance, Friends and Marriage, and Loneliness sections. The remaining portions of the ADOS were not transcribed. Within each activity, conversational turns were segmented into individual utterances (or “C-units”), each consisting of (at most) a main clause and any subordinate clauses modifying it.

**Measures Derived from ADOS Transcripts.** ADOS transcripts were used to compute overall MLUM [Brown 1973] using SALT software [Miller and Chapman, 1985]. MLUM is a simple, face-valid measure of morphological and syntactic complexity recommended as a benchmark of spoken language development in children with autism [e.g., Tager-Flusberg et al., 2009]. These transcripts were also used to count *uh*s, *um*s, and fluent words (words which are not part of a maze) for each participant. In the case that a child produced multiple consecutive fillers within a single utterance (e.g., she *had the um um starfish*), only the first filler was counted. Immediately-repeated fillers were excluded on the hypothesis that a repeated filler is not statistically independent of the preceding filler, and thus their inclusion would violate the independence assumptions of the quantitative analyses. The full data set contained 1,261 tokens of *uh* and 2,523 tokens of *um*.

---

1 An anonymous reviewer asks whether data gathered from separate ADOS modules can be compared. To test whether our results were sensitive to this, we repeated all statistical analyses excluding data from children who completed the ADOS Module 2. The results indicated that excluding data from children who completed Module 2 had no effect on the results obtained.

2 An anonymous reviewer asks whether the exclusion of immediately-repeated fillers may have influenced the results. To test this, we repeated all statistical analyses without these exclusions and found that it had no effect on the results obtained.
Filler annotation quality was assessed retrospectively using a stratified random sample of the full dataset. The sample contained four utterances per child, two of which had been transcribed as containing one or more fillers (uh or um) and two of which had been transcribed without any fillers. Utterances were excluded from this sample if they contained unintelligible words or if the examiner’s speech overlapped the child’s speech. Audio files of these sample utterances were extracted with a two-second fade-in/fade-out. These files were then transcribed, according to the same guidelines, by two experienced transcribers, neither of whom participated in the initial transcription efforts. Both transcribers were blind to participant identity and group assignment.

When original and retrospective annotations both contained a filler, the original and retrospective annotators transcribed the same filler type (uh or um) in 95% and 91% of cases, with Cohen’s kappa (κ) of 0.893 and 0.810, respectively (see Table 1). This corresponds to “almost perfect” agreement according to Landis and Koch’s [1977, p. 165] qualitative guidelines. These retrospective transcriptions were also used to assess agreement for the presence or absence of fillers, ignoring filler type. In 87% of cases, the original and retrospective transcriptions agreed on presence or absence of fillers, with κ of 0.737 and 0.729, respectively, indicating “substantial” agreement according to the Landis and Koch guidelines.

### Statistical Analysis

Inferential analyses were conducted using mixed effects logistic regression [Pinheiro and Bates, 2000] with a per-subject random intercept. Compared to conventional (i.e., fixed effects) logistic regression, this method provides a principled solution to problems of non-independence and heteroscedascity arising when subjects contribute different numbers of observations, as is the case here. The primary independent variable was participant group (ASD, SLI, or TD). All models also included three participant-linked covariates: chronological age, full-scale IQ, and MLUM. Each token was coded for ADOS activity (Play, Description of a Picture, Telling a Story from a Book, or Conversation). A binary predictor was used to code whether a token was utterance-initial or noninitial. To facilitate interpretation, continuous variables were z-transformed, and sum coding was used to encode categorical variables. The log-likelihood ratio test was used to test for significance of individual predictors, and the Tukey HSD test was used to test for significant differences within factor groups. Exploratory analyses were conducted by measuring correlations with Kendall’s τb, a nonparametric correlation statistic.

### Results

#### Group Characteristics

Summary statistics for the three diagnostic groups are reported in Table 2.

#### Inferential Analyses

Three separate inferential analyses were performed. The first two compared children’s productions of uh and um, respectively, to their productions of fluent words,
Mixed effects logistic regression on \(uh\) rate; predictors which are positively associated with \(uh\) have positive log-odds and predictors which are negatively associated with \(uh\) have negative log-odds. CA, chronological age in years; FSIQ, full-scale IQ.

Mixed effects logistic regression comparing \(uh\) and \(um\) frequencies; predictors which favor \(um\) have positive log-odds and predictors which favor \(uh\) have negative log-odds. CA, chronological age in years; FSIQ, full-scale IQ.

### Table 3. Results for Regression on \(uh\) Rate

<table>
<thead>
<tr>
<th></th>
<th>Log-odds</th>
<th>s.e.</th>
<th>(\chi^2)</th>
<th>(P(\chi^2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-5.462</td>
<td>0.11</td>
<td>2.27</td>
<td>0.306</td>
</tr>
<tr>
<td>Group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASD</td>
<td>0.191</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLI</td>
<td>-0.208</td>
<td>0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD</td>
<td>0.017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>0.000</td>
<td>0.09</td>
<td>0.00</td>
<td>0.986</td>
</tr>
<tr>
<td>FSIQ</td>
<td>0.089</td>
<td>0.12</td>
<td>0.51</td>
<td>0.474</td>
</tr>
<tr>
<td>ADOS Activity:</td>
<td></td>
<td></td>
<td>52.67</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Play</td>
<td>0.047</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description of a picture</td>
<td>0.029</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telling a story from a book</td>
<td>-0.382</td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conversation</td>
<td>0.306</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context:</td>
<td></td>
<td></td>
<td>975.18</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Initial</td>
<td>0.915</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noninitial</td>
<td>-0.915</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mixed effects logistic regression on \(um\) rate; predictors which are positively associated with \(um\) have positive log-odds and predictors which are negatively associated with \(um\) have negative log-odds. CA, chronological age in years; FSIQ, full-scale IQ.

### Table 4. Results for Regression on \(um\) Rate

<table>
<thead>
<tr>
<th></th>
<th>Log-odds</th>
<th>s.e.</th>
<th>(\chi^2)</th>
<th>(P(\chi^2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-4.999</td>
<td>0.12</td>
<td>16.13</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASD</td>
<td>-0.588</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLI</td>
<td>0.042</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD</td>
<td>0.546</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>0.103</td>
<td>0.11</td>
<td>0.82</td>
<td>0.364</td>
</tr>
<tr>
<td>FSIQ</td>
<td>-0.054</td>
<td>0.15</td>
<td>0.13</td>
<td>0.719</td>
</tr>
<tr>
<td>ADOS Activity:</td>
<td></td>
<td></td>
<td>216.76</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Play</td>
<td>-0.215</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description of a picture</td>
<td>0.265</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telling a story from a book</td>
<td>-0.438</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conversation</td>
<td>0.388</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context:</td>
<td></td>
<td></td>
<td>1630.39</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Initial</td>
<td>0.848</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noninitial</td>
<td>-0.848</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5. Results for Regression on \(uh\) vs. \(um\)

<table>
<thead>
<tr>
<th></th>
<th>Log-odds</th>
<th>s.e.</th>
<th>(\chi^2)</th>
<th>(P(\chi^2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(intercept)</td>
<td>0.433</td>
<td>0.16</td>
<td>16.29</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASD</td>
<td>-0.795</td>
<td>0.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLI</td>
<td>0.263</td>
<td>0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD</td>
<td>0.531</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>0.124</td>
<td>0.14</td>
<td>0.75</td>
<td>0.387</td>
</tr>
<tr>
<td>FSIQ</td>
<td>0.075</td>
<td>0.19</td>
<td>0.16</td>
<td>0.691</td>
</tr>
<tr>
<td>ADOS Activity:</td>
<td></td>
<td></td>
<td>24.32</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Play</td>
<td>-0.262</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description of a picture</td>
<td>0.160</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telling a story from a book</td>
<td>-0.126</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conversation</td>
<td>0.229</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context:</td>
<td></td>
<td></td>
<td>1.93</td>
<td>0.164</td>
</tr>
<tr>
<td>Initial</td>
<td>0.064</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noninitial</td>
<td>-0.064</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

i.e., those not part of a maze; these test for group differences in the use of these two fillers while controlling for any group differences in fluent verbal output. The third analysis compares children’s productions of \(uh\) to productions of \(um\), controlling for any group difference in overall filler production.

**Uh rate.** In the first mixed effects regression, each token of \(uh\) was coded as a “hit” and each fluent word (those not part of a maze) as a “miss.” The results are shown in Table 3. There was no effect of group, chronological age, or full-scale IQ. However, there was a significant effect of ADOS activity \((\chi^2 = 52.67, P < 0.001)\), and post hoc tests identified significant contrasts between nearly all pairs of ADOS activities, with the highest rate of \(uh\) occurring during the conversation activity (Telling a Story from a Book < Description of a Picture = Description of a Picture < Conversation, all \(P < 0.001)\). Tokens of \(uh\) were also more likely in utterance-initial position than in non-initial position \((\chi^2 = 975.18, P < 0.001)\).

**Um rate.** The second mixed effects regression investigated \(um\) use. Each token of \(um\) was coded as a “hit” and each fluent word as a “miss.” The results are shown in Table 4. There was a main effect of group \((\chi^2 = 16.13, P < 0.001)\). Post hoc tests revealed that the ASD group used \(um\) at a significantly lower rate than children with typical development \((P < 0.001)\); differences between the other pairs of groups were nonsignificant. Chronological age, full-scale IQ, and MLUM were not associated with use of \(um\). Once again, there were significant effects of ADOS activity \((\chi^2 = 216.76, P < 0.001)\) and post hoc tests identified significant contrasts between several pairs of ADOS activities (Telling a Story from a Book < Play < Description of a Picture < Conversation, all \(P < 0.001)\). Tokens of \(um\) were more likely in utterance-initial position than in non-initial position \((\chi^2 = 1630.39, P < 0.001)\).

**Uh vs. um use.** The final regression examined filler choice by comparing \(uh\) and \(um\) frequencies. Tokens of \(um\) were coded as “hits” and tokens of \(uh\) as “misses,” thus controlling for any group differences in overall filler

---

6 Gorman et al./Uh/um in children with ASD/language impairment

 INSAR
Kendall’s *s* for executive function, language, and social ability, as measured by ADOS SA, ADOS-G social affect calibrated severity score (SCQ CTS), a parent-reported measure. This effect was reliable across the three groups, with weak negative correlations between language ability and social function. We computed correlation coefficients for the association between um–uh ratio and each of these measures; separate analyses were conducted for each of the three groups. Within each group, we found robust group differences in Filler use compared to play, conversation strongly favored um (P = 0.001). The um–uh comparison is depicted in Figure 1. Each dot represents the percentage of fillers which are um—i.e., um/(uh + ums)—for each participant.

### Associations Between Filler Use and Other Measures

We also conducted an exploratory analysis to investigate whether within-group heterogeneity in filler use was associated with chronological age, intellectual ability, executive function, structural and pragmatic language, or social communication. We computed correlation coefficients for the association between the um–uh ratio and each of these measures; separate analyses were conducted for each of the three groups. Within each group, P-values were adjusted for false discovery rate [Benjamini and Hochberg, 1995]. Many of these tests are complementary—i.e., several pairs of independent variables are highly correlated and measure closely-related constructs—and the resulting statistical tests are very likely underpowered (particularly in the SLI group), so even the adjusted P-values should be interpreted with caution.

The results are shown in Table 6. There were no reliable associations between um–uh ratio and chronological age, intelligence, or executive function. There was a significant association between um–uh ratio and MLUM in the TD group (r = 0.34, P = 0.020). There were no reliable associations between um–uh ratio and scores on the CELF, the CCC-2, or the ADOS SA. In all three groups, there were weak negative correlations between um–uh ratio and the SCQ communication total score (SCQ CTS), a parent-reported measure. This effect was marginal in ASD (r = −0.29, P = 0.073) and nonsignificant in TD and SLI. (Note that higher scores on the SCQ CTS and ADOS SA indicate greater degrees of impairment.)

### Discussion

In this study, we investigated the use of uh and um in a sample consisting of children with ASD, SLI, and typical development. These fillers play a subtle but important role in everyday life and atypical use of fillers by speakers with ASD may contribute to difficulties engaging in conversations with others. Although we did not find any group differences in uh rate, we found robust group differences in um rate and in um–uh ratio. Approximately 40% of the fillers used by children with ASD were um, but um accounted for more than 70% of the fillers used by children in the TD and SLI groups.
Participants in this study were all highly verbal. However, there was considerable variability in their structural language abilities. Half of the 50 participants with ASD had a CELF CLS more than one standard deviation below the normative mean, as did all participants in the SLI group. On three standard measures of language—MLUM, verbal IQ, and the CCC-2 GCC—the ASD and SLI groups were well-matched, and both groups had significantly lower mean values than the TD group (see Table 2). We found that the SLI and TD groups both had a significantly higher \textit{um}-\textit{uh} ratio than the ASD group, and were not significantly different from each other. This suggests that a low use of \textit{um} is specific to ASD, but independent of impairments in structural language, which are relatively prevalent among—but not specific to—children with ASD [e.g., Leyfer et al. 2008; Loucas et al. 2008]. The possibility of an autism-specific pragmatic deficit is particularly interesting in light of findings suggesting that pragmatic difficulties are also common in children with SLI [e.g., Bishop, 2001].

We hypothesized that different ADOS activities might influence filler use and choice. In this study, ADOS activity emerged as one of the most robust predictors of filler rates. To take an extreme example, \textit{um} was more than twice as common during the conversation activity than during the Telling a Story from a Book activity, an effect that is nearly as large as differences between the ASD and TD groups (see Table 4). As mentioned earlier, studies of typical adults suggest that cognitive load contributes to elevated disfluency rates. Our results are consistent with this, under the assumption that social-emotional conversation is particularly cognitively loading [cf. Adams et al., 2002], and that the demands of face-to-face discussion of emotional topics limit speakers’ abilities to effectively plan and monitor their speech. More generally, this result highlights the importance of controlling for topic in quantitative studies of pragmatic language.

This study had several limitations. First, participants were drawn from a relatively wide age range (4–8). Although chronological age was included as a covariate in regression analyses, developmental differences may have obscured important group differences. Second, the majority of the participants were male. Consequently, we lacked statistical power to investigate gender differences, although there is some evidence that typical adult female speakers produce more \textit{ums} than their male peers [Acton, 2011; Tottie, 2011]. Furthermore, we did not investigate the role of socioeconomic status or ethnicity, although class and ethnicity may play a role in pragmatic language, including use of \textit{uh} and \textit{um} [Rayson, Leech, & Hodges, 1997; Tottie, 2011]. Another limitation is that the diagnostic groups were defined using strict cutoffs for SLI and ASD; different cutoffs might produce different results. Finally, all participants were high-functioning, limiting the generalizability of these results to the larger population of individuals with ASD.

The current study was limited to the English fillers \textit{uh} and \textit{um}. However, the general patterns documented here are not necessarily limited to children acquiring English. Fillers appear to be a linguistic universal [Allwood et al. 1990, p. 33], and virtually all languages have at least two distinct fillers [Clark and Fox Tree, 2002; Wilkins, 1992]. As in English, different fillers tend to exploit different discourse niches. In Dutch, for instance, \textit{ah} [a] favors phrase-medial position and tends to precede shorter pauses, while \textit{um} [am] favors phrase-initial position and tends to introduce longer pauses [Swerts, 1998, p. 490], just as in English. Similarly, in Japanese, the fillers \textit{ato} and \textit{sono} pattern with English \textit{uh}, whereas \textit{e} and \textit{eto} pattern with English \textit{um} [Watanabe and Ishi, 2000; Watanabe, 2001, 2002]. Thus, it is possible that similar patterns will be found in children acquiring other languages, although we leave this as a topic for future research.

One potential confound in this study concerns the role that prosody plays in the perception of fillers. In the speech of typical adults, there are distinct prosodic cues associated with \textit{uh} and \textit{um} [Levelt and Cutler, 1983; Shriberg and Lickley, 1993]. Listeners can use these cues to identify fillers, even in unfamiliar languages [Lai et al., 2007] or in speech that has been low-pass filtered [Lickley, Shillcock, & Bard, 1991]. Since many children with ASD are thought to exhibit atypical speech prosody [e.g., Paul et al., 2005; Peppé et al., 2007; van Santen et al., 2010], this might make it more difficult for transcribers—who presumably rely on these prosodic cues when making transcriptions—to detect fillers in these children. It remains to be seen whether disfluencies are associated with different prosodic cues in the speech of children with ASD. However, there were no group differences in overall rates of fillers or annotator reliability, suggesting that atypical prosody in children with ASD cannot fully account for the group differences we found.

It is not known whether individual abilities in production and perception of fillers are associated, but it is possible that in addition to atypical filler use, children with ASD may also have difficulties perceiving fillers (i.e., due to difficulties with prosody perception) or interpreting the full set of social cues that fillers convey. While our data does not directly address speech perception, this may be another fruitful topic for future work.

Our exploratory analysis suggests that \textit{um}-\textit{uh} ratio is associated with a parent-endorsed measure of social communication ability. Thus, a child’s filler use may influence a parent’s assessment of his or her child’s
communicative competence, but further research is needed to investigate parental perceptions of communicative abilities in ASD to determine the extent of this effect.

Our inferential analyses uncovered robust differences between children with and without ASD. Given that social-communicative deficits are a defining feature of ASD, our findings provide convergent evidence for the essentially social function of fillers in the speech of typical individuals [e.g., Brennan and Schober, 2001; Clark, 1994; Fox Tree, 2001]. Our findings also contribute to our understanding of the inherent difference between *uh* and *um*. Similar findings are reported in a recent study by Irvine [2014], who also found that children and adolescents with ASD produce a lower *um* rate than typical controls. In that study, a third group of participants, who received a prior diagnosis of ASD but who no longer meet diagnostic criteria, produced *um* at similar rates to TD peers and at significantly higher rates than peers with ASD. If group differences in *um* use and correlations between *um* use and social communication abilities are replicated, fillers, along with other subtle aspects of pragmatic language, may be a useful target for intervention, particularly in individuals with ASD who are verbal and high-functioning.

**Conclusions**

We have shown that children with high-functioning ASD produce *um* at a rate much below that of children without ASD, and that use of *um* is positively associated with parents’ assessments of their child's social communication abilities. Crucially, use of *um* is not associated with language impairment within the group of children diagnosed with ASD, and children with SLI use *um* at a similar rate as their TD peers. Filler use is an easily quantified feature of pragmatic language that, in concert with other behavioral features, may distinguish ASD and SLI, a notoriously difficult differential diagnosis [Bishop and Norbury, 2002; Bishop et al., 2008; Cox et al., 1999]. We believe that this general approach—using contextualized natural language samples to quantify features of pragmatic language—will provide much-needed conceptual precision and complement existing methods of diagnosis and phenotypic characterization based on clinical observation or structured assessment.

**Acknowledgments**

We thank Mabel Rice for helpful discussion on criteria for classifying specific language impairment, and Lauren Kenworthy for assistance with measures of executive function. Thanks also to Julianne Myers, Hilary Prichard, Emily Prud’hommeaux, and audiences at IMFAR 2014 for their comments on this work. We also thank the two anonymous reviewers. L. Olson is now with the Marcus Autism Center at Emory University. This material is based on work supported by the National Institute on Deafness and Other Communication Disorders of the National Institute of Health under awards R01DC007129 and R01DC012033, and by Autism Speaks under Innovative Technology for Autism Grant 2407. The content is solely the responsibility of the authors and does not necessarily represent the official views of the granting agencies or any other individual.

**References**


