

NORTHEASTERN UNIVERSITY  
COLLEGE OF SCIENCE

**THE EFFECT OF PRESCRIPTIVE RULES AND INSTRUCTIONS ON  
THE GRAMMATICALITY JUDGMENT TASK**

BY

**PARKER T. ROBBINS**

An Undergraduate Honors Project Submitted to the Linguistics Program  
in partial fulfillment of the requirements for  
Distinction in the Major

Heather Littlefield, Ph.D., Project Advisor  
Adam Cooper, Ph.D., Second Reader

May 2019

## ABSTRACT

### THE EFFECT OF PRESCRIPTIVE RULES AND INSTRUCTIONS ON THE GRAMMATICALITY JUDGMENT TASK

Author: Parker T. Robbins, B.S. in Linguistics  
Project Advisor: Heather Littlefield, Ph.D., Linguistics Program  
Second Reader: Adam Cooper, Ph.D., Linguistics Program  
Department: Linguistics Program, College of Science  
Year: 2019

The grammaticality judgment task (GJT) attempts to understand language phenomena through speaker intuition, which requires speakers' judgments to reflect implicit linguistic knowledge and not rules prescribing "correct" language use. The validity of the GJT has been investigated before (Schütze 2016), but only one study, Cowart (1997), has investigated whether task instructions affect participants use of prescriptive rules. This paper presents a follow-up study with 200 participants, new stimuli, and adapted versions of Cowart's instructions, two of which, INTUITIVE PLAIN and INTUITIVE SCHOOL, target intuitive knowledge and two of which, PROFESSOR and TUTOR, simulate the role of an instructor. Both intuitive sets ask participants for "gut reactions," but only INTUITIVE SCHOOL tells them explicitly not to use "school grammar." The PROFESSOR condition has participants imagine they are English professors and the TUTOR condition has them imagine they are tutoring a friend learning English. Stimuli test whether participants prefer sentences following a prescriptive rule when either is grammatical and when the application of a prescriptive rule would make a sentence ungrammatical. As English professors are associated with using "correct" English, I hypothesized that participants in the PROFESSOR condition would prefer prescriptively correct sentences compared to the baseline, INTUITIVE PLAIN. I found that while PROFESSOR condition participants were more likely to respond prescriptively when both options were grammatical, TUTOR condition participants were more likely to choose an ungrammatical prescriptive response. Overall, this study suggests more care should be taken in designing GJT instructions, especially when prescriptive rules may not coincide with speakers intuitions.

**KEYWORDS:** Grammaticality judgment task, methods, psycholinguistics, prescriptivism, syntax.

# The Effect of Prescriptive Rules and Instructions on the Grammaticality Judgment Task

Parker T. Robbins

## 1 Introduction<sup>1</sup>

The grammaticality judgment task is a fundamental research method in modern linguistics. It attempts to access speakers tacit knowledge about language by asking them to judge how well-formed—or natural—an utterance is. These judgments can be used to discover syntactic facts about specific languages and to build a theory of language as a whole. Questions about the empirical validity of this method stretch at least as far back as Labov (1975), and its use continues to be questioned more recently. Riemer, for example, challenges the “widespread assumption” that grammaticality judgments constitute acceptable evidence for informing linguistic theories because of their instability, which can feature even within individual participants over time (2009: 614). This study aims to contribute to the empirical validation of the grammaticality judgment task by investigating whether judgment study participants are more likely to respond according to prescriptive grammar rules given different types of instructions. After outlining previous research on the grammaticality judgment task, I explain the present study, methods, and results. Finally, I discuss the results in the context of previous research and make suggestions for future research on the empirical validity of the grammaticality judgment task.

## 2 Background

### 2.1 Previous research on the grammaticality judgment task

There is a body of research that has examined the role of various task- and participant-related factors, as Schütze (2016) outlined in his review of the topic. A few examples illustrate the types of effects that have been investigated in the course of empirically validating this task. Snyder (2000), for example, found evidence for syntactic satiation, a phenomenon in which previously unacceptable utterances become increasingly acceptable with exposure, suggesting that this effect could make some judgments unreliable. Bader and Haeussler (2010), on the other hand, found that different elicitation techniques, including untimed binary grammaticality judgment tasks, speeded binary grammaticality judgment tasks, and a type of gradient grammaticality judgment task called magnitude estimation, were relatively stable, indicating that the specific type of judgment methodology might not be as important as previously thought. Finally, Smith (2011) found that auditory distraction has little effect on accuracy and mixed effects on processing speed in grammaticality judgment tasks. These examples show that, although further research is needed, there are high-quality recent psycholinguistic studies that attempt to empirically validate the grammaticality judgment task from several methodological points of view.

Not all aspects of this method have been studied, however, and its reliability unfortunately continues to be taken for granted by many theoretical linguists. The continued empirical validation

---

<sup>1</sup>This work was funded by an Undergraduate Research and Creative Endeavor Award from Northeastern University’s Office of Undergraduate Research and Fellowships. My project would not have been possible without them or Northeastern’s IRB, who approved the study (# 18-12-20). I would like to thank Professor Littlefield for all of her help designing and implementing this study and I would like to thank Professor Cooper for his insightful comments and advice about my project. I would like to thank Jayden Ziegler and Anthony Yacovone of the Snedeker Lab at Harvard for their advice as well. Finally, I would like to thank Harvard’s statistical consulting services for help designing my analysis and interpreting my results. All code and data used in this study can be found at <https://github.com/parkerrobbins/2019-judgment-study/>.

of the grammaticality judgment task is crucial to ensure that linguists are collecting reliable judgments from participants which accurately reflect speakers implicit knowledge about language. In order to build scientific theories of language, the methods used to collect linguistic data need to be empirically grounded.

## 2.2 Examining the role of prescriptive rules on grammaticality judgments

As Cowart (1997) observed, early researchers who were concerned about the validity of the grammaticality judgment task questioned whether linguists could even access speakers implicit linguistic knowledge without interference from extralinguistic factors. These external factors might include explicit knowledge about language or attitudes toward nonstandard or stigmatized language varieties. The assumption of many researchers is that the linguistic intuition being accessed in grammaticality judgment tasks is primarily implicit, accessing the deeper psychological basis of a speakers linguistic system. Love (1989), however, proposed that this assumption may not be warranted, as the psychological underpinnings of linguistic knowledge are not easily separable from social attitudes about language. Some researchers, such as Valian (1982), have suggested that training in linguistics is necessary for judgment study participants, as only those with this training can be assumed to reliably distinguish intuitive and explicit language knowledge. One might expect that knowledge of linguistic concepts and investment in particular hypotheses about language may cause language researchers to respond differently than linguistically naïve participants. Studies such as Spencer (1973), Ross (1979), and Dąbrowska (2010) indicate that the ratings of participants with training in linguistics—ranging from undergraduate to doctoral level—do tend to be systematically different from those of non-linguists. This suggests that knowledge about linguistics biases participant responses, rendering the solution proposed by Valian unsatisfying. It is therefore important to investigate whether participants use explicit knowledge of and attitudes about language in their judgments, and how that may affect these judgments, especially when these factors might distort the intuitions that linguists are attempting to capture. Prescriptive rules, precepts taught in schools designed to teach students how to speak “correctly,” represent one phenomenon that can help to answer this question. One familiar example of a prescriptive rule is the ban against splitting infinitives. Consider the following two sentences in (1).

- (1) a. I am going to quickly go to the store.  
 b. I am going to go quickly to the store.

While both (1b) and (1a) are generally considered natural linguistic constructions in English, albeit with slightly different meanings, (1a) is not considered prescriptively correct because the adverb *quickly* comes between *to* and *go*, thereby “splitting” the infinitive *to go*. As both constructions are natural in English but only one is acceptable based on prescriptive standards, this is an example of explicit language knowledge interfering with implicit linguistic knowledge. In other words, there is a conflict between tacit linguistic knowledge that speakers use to construct sentences in everyday life and the prescriptions taught by educators in schools. When linguists are eliciting grammaticality judgments, they are generally interested in describing language to understand how it works in the mind of a speaker. Speakers’ perceptions of “correct language” that come from explicit instructions may thus obscure the implicit knowledge that a linguist is attempting to capture.

Several studies have examined the use of prescriptive rules in grammaticality judgments, although compared to other factors that have been studied, research in this area remains limited.

Schmidt and McCreary (1977) investigated the effect of three prescriptive rules on the judgments of native and non-native English speakers. These prescriptive rules included agreement with *there is* and *there are* (see (2a) for an example), the use of *their* as a third-person singular pronoun (2b), and the exclusive use of *I* as a subject and *me* as an object (2c). In these examples, the choice on the right side of the slash indicates strict adherence to prescriptive rules.

- (2) a. There is/are two minutes remaining on the clock.
- b. Each student should hand in their/his or her homework on time.
- c. You and me/I have been best friends since kindergarten.

Schmidt and McCreary selected participants with varying levels of education, including sixth graders, twelfth graders, undergraduate students, and graduate students. First, these participants completed a production task in which they were asked to transform sentences based on a clue, which would give participants the option to apply a prescriptive rule or ignore it. The sentences in (3) give an example of the original sentence provided (3a), the clue directing participants to transform that sentence (3b), and potential responses (3c).

- (3) a. My sister and I received the letter.
  - b. The letter was received...
  - c. The letter was received by my sister and I/me/myself.
- (Schmidt and McCreary 1977: 424)

Participants then completed two judgment surveys, which both listed series of three semantically similar but syntactically different sentences. In these surveys, participants indicated which form they would use more often and then what the “correct” form would be. Overall, Schmidt and McCreary found that students with higher levels of education were more likely to prefer, use, and report using prescriptively correct forms. As many participants in grammaticality judgment tasks conducted at universities are undergraduate students, it is reasonable to expect that prescriptive rules may affect judgments in some of those studies, as well.

Bradac et al. (1980) investigated whether participants with and without training in linguistics rate sentences using prescriptive grammar rules. These participants completed a grammaticality judgment task, and the authors found that participants without linguistic training (and therefore likely unfamiliar with the distinction between linguistic intuition and prescriptive knowledge) do not always use prescriptive rules to make judgments about grammaticality, and that some rules such as the ban on double negatives are more salient than others, such as number agreement with mass and count nouns. This indicates that the influence of a given prescriptive rule may vary in a judgment task, making it more difficult to control or account for this effect within and across participants.

Finally, Henry (2005) examined the role of prescriptivism in gathering judgments from speakers of nonstandard dialects. When forms in a nonstandard dialect violate prescriptive rules in that community's standard dialect, speakers may, she suggested, provide judgments that adhere to prescriptive rules due to attitudes about their dialect and group pressure. The author offered some suggestions for mitigating these effects, including testing participants separately and presenting questions orally. Unfortunately, she did not present her data or analyze them in a systematic way, so the bulk of the evidence she presents remains anecdotal.

Though the results of these studies are not uniform, they generally find that prescriptivism can affect how participants perform in grammaticality judgment tasks. Moreover, as there is not a large body of work on this subject, the role of prescriptive rules in speakers judgments requires further investigation.

### 2.3 Task instructions and prescriptive rules

The results of Schmidt and McCreary (1977) suggest a correlation between level of education and knowledge and use of prescriptive rules. Intuitively, this makes sense, as formal education is a major source of prescriptive rules and other explicit knowledge of language. It is therefore reasonable to imagine that how rules frame a task may affect whether participants rely on explicit knowledge of prescriptive rules or whether they report their “gut reaction” to a particular utterance. There appears to be only one study in the literature that investigates whether instruction type can prime prescriptive judgments and thus affect the overall results in a judgment study.

In chapters one and four of Cowart (1997), the author presents an informal study that attempts to address this question. Participants were divided into two groups, one that received intuitive instructions and one that received prescriptive instructions. The intuitive instructions reflect instructions frequently given for tasks, requesting participants to use “gut knowledge” and not “rules [they] may have learned about ‘proper’ or ‘correct’ English” (57). The prescriptive instructions, on the other hand, asked participants to imagine a professor in a high-level undergraduate English course and judge each sentence based on whether they thought this professor would accept it. As English courses typically feature prescriptive grammar rules, Cowart hypothesized that if participants did vary in the extent to which they applied prescriptive knowledge to judgments, this application would be higher in the latter condition. He did not find any statistically significant differences between the two groups, though as he noted, that does not indicate that no such effect of instruction type exists in general. Cowart’s relatively small sample size of 43 and choice of a four-point Likert-type scale (rather than a forced-choice task, for example) may make it more difficult to detect a small effect between instruction conditions if one exists. In designing the study, Cowart had strong predictions for the acceptability of sentences presented, as the example stimuli in (4) illustrate.

- (4) a. Cathy’s parents require that Paul support himself.  
 b. \*Paul requires that Cathy’s parents support himself.  
 (Cowart 1997: 20)

Despite his predictions about the grammaticality of (4a) and the ungrammaticality of (4b), Cowart did not actually vary sentence stimuli based on a widely-known prescriptive rule or contrast what is generally considered acceptable with what is prescriptively considered “correct,” making it unclear how to tell under any conditions whether differences between the groups were a result of participants applying a prescriptive rule or not.

Cowart explained that most previous studies had ignored how instructions frame the task, preferring to use intuitive instructions that appeal to the gut reactions and allow participants to respond “on any grounds (other than school grammar) that seem appropriate to the informant” (56). He viewed this approach as a reflection of the widespread belief that participants cannot control what type of judgments they provide, prescriptive or not, and thus that more emphasis ought to be placed on “experimental control rather than direct manipulation of the informants approach to the

task” (56). This sets up a false dichotomy, however: if instructions, even intuitive instructions, are not presented in a standard way, this *is* an issue of experimental control, regardless of whether the instructions explicitly guide the criteria participants use to apply judgments or whether they subtly and subconsciously evoke particular language attitudes. Further, to my knowledge there is no evidence that most participants without training in linguistics are familiar enough with the distinction between grammatical intuition and explicit language knowledge to differentiate these even when explicitly asked to, as in the intuitive instructions.

As Schütze and Sprouse (2014) point out, the lack of further investigation on this question can be attributed to a widespread belief among researchers that instruction type does not affect participant judgments. However, previous research on the link between prescriptive rules and linguistic judgments suggests that if instructions, below the level of consciousness or not, can prime participants to access particular language attitudes, this would have implications for many studies, especially those which investigate nonstandard dialects or that seek to investigate a phenomenon that has a different status in a speakers intuition than what is considered “correct” prescriptively.

This work is an essential step on the path to fully validating the grammaticality judgment task, ensuring that linguists are gathering high-quality data when using this method to build linguistic theories. Possible applications of an improved grammaticality judgment task include better documentation of endangered languages, a task which frequently involves nonstandard or stigmatized language varieties, or creating linguistically-savvy artificial intelligence, which requires detailed and accurate input to function correctly. Validating the grammaticality judgment task is thus crucial not only for improving the scientific study of language, but also for the real-world applications that such a study can inform.

#### **2.4 The present study**

The purpose of this study is thus to reproduce the study presented in Cowart (1997) and expand it by adding more instruction conditions, using a phenomenon that contrasts intuition and prescriptive attitudes, and recruiting a much larger sample size than was done previously. These improvements facilitate the detection of a small effect and permit an unambiguous interpretation of how participants use instructions. There were two main hypotheses: (a) that participants in the PROFESSOR condition were more likely to respond prescriptively when presented with two grammatical sentences and (b) that participants in the PROFESSOR condition were more likely to respond prescriptively when such a choice would be ungrammatical.

The study was also exploratory in nature: A TUTOR instruction condition was added, drawn from a suggestion in Cowart (1997) that was not actually tested in his study. This condition was suggested by Cowart as a way to direct participants attention to an acceptability task in a way that seems natural but does not explicitly call on prescriptive judgments. The claim that TUTOR instructions may provide a middle ground between PROFESSOR and INTUITIVE instructions was also tested. Finally, Cowart’s INTUITIVE instructions explicitly asked participants to ignore “school grammar.” As this may have a counterproductive effect, effectively priming participants to use this knowledge, I created a second version of these INTUITIVE instructions that removes this potential bias to test whether this changed participant responses.

### **3 Methodology**

#### **3.1 Participants**

Participants were 200 workers on Amazon Mechanical Turk, a platform on which requesters can compensate anonymous workers for small tasks. The participants (86 female, 113 male, and one other; mean age = 36.69, SD = 10.75, one value unreported) were paid \$4.00 for their participation. This platform has been validated for linguistic research in general (Schnoebelen and Kuperman 2010) and for acceptability judgments in particular (Sprouse 2011). I recruited workers who have a 95 percent acceptance rate on past tasks as a first measure to ensure data quality (Peer et al. 2013). Additionally, participants were removed from the study (and not compensated, as is standard) if they did not pass an instruction comprehension question or either of two attention checks in the survey. Participants were asked to participate only if one of their native languages was English, though to discourage deception any participants who did not report English as a native language or whose reported native languages did not include English were excluded but still compensated. Finally, I excluded (but still compensated) participants who had reported having taken a linguistics course, as these participants may be less naïve to the prescriptive rule manipulation in the task. Data collection continued until acceptable data from 200 participants had been collected.

#### **3.2 Materials**

The grammaticality judgment task was deployed using the Ibex Farm platform, a free server for running internet-based language experiments using the Ibex software (Drummond 2013). Further, I used the PennController plugin for this software, which allows for more precise and user-friendly control of elements in an Ibex survey, including the ability to create templates that generate trials based on stimuli in a spreadsheet (Zehr and Schwarz 2018). Participants were randomly assigned to each instruction condition, and all questions after the instruction comprehension check were presented in a pseudorandom order such that no more than two critical sentence items appeared between each filler item and that the attention check questions were distributed evenly among the sentence pairs. Additionally, the presentation order of individual choices in sentence pairs were random.

##### **3.2.1 Task instruction conditions**

Three instruction conditions were created based on instruction types presented by Cowart. The first two, the INTUITIVE and PROFESSOR conditions, were directly adapted from Cowart's study, with the appropriate changes made to reflect a forced-choice task rather than a Likert-type scale rating task (1997: 57). The PROFESSOR condition was designed to elicit prescriptive judgments, while the INTUITIVE condition was designed to evoke intuitive responses that linguists are frequently seeking to collect. The INTUITIVE SCHOOL instructions and the INTUITIVE PLAIN instructions were the same as Cowart's intuitive instructions except that the latter did not include a reference to "school grammar" at all. The TUTOR instructions ask participants to imagine they are tutoring a friend who was learning English, and to judge whether a given sentence sounds like it was produced by a native speaker (1997: 91). The full text of the instructions can be found in appendix A.

##### **3.2.2 Sentence pair stimuli**

The rule against stranded, or sentence-final, prepositions dates from the eighteenth century, but continues to be reinforced through the Microsoft Word grammar checker (Curzan 2014). As this is a well-established prescriptive rule that remains relevant today, it was chosen as the basis for



creating sentence stimuli. Three types of sentence pairs were created: 20 with prepositional phrasal verbs (PPV; e.g., *compare to*), 20 with particle-prepositional phrasal verbs (PPPV; e.g., *put up with*), and 20 fillers. The PPV sentence pairs made it possible to test whether the participant prefers the construction that follows the prescriptive rule when both sentences in the pair are grammatical, as (5) illustrates.

- (5) a. The master violinist was a musician whom few professionals could really compare to.  
b. The master violinist was a musician to whom few professionals could really compare.

While violations of this prescriptive rule involving prepositional phrasal verbs are addressed by pied-piping, the phenomenon by which a preposition is raised to the position before the complementizer of an embedded clause, pied-piping is not possible for both the particle and preposition of a particle-prepositional phrasal verb (Heck 2008). The PPPV sentence pairs, an example of which is illustrated in (6a) and (6b), made it possible to test whether a participant applies the prescriptive rule even when doing so would create an ungrammatical sentence. Crucially, as (6c) shows, particle-prepositional phrasal verbs can show more than one type of movement. However, because the objective of these sentence pairs is to investigate the competition between prescriptive rules and linguistic intuition, it is not expected that this additional possibility affected the forced choices that participants made.

- (6) a. This kind woman was a friend whom my aunt would frequently catch up with.  
b. \*This kind woman was a friend up with whom my aunt would frequently catch.  
c. This kind woman was a friend with whom my aunt would frequently catch up.

Finally, fillers are a standard element of grammaticality judgment tasks which serve to distract participants from the experimental manipulation. In this case, I varied the placement of prepositional phrases in the sentence; in half of the sentences, both versions were grammatical (7) and in the second case, one of the sentences is less acceptable (8). The full set of all types of sentence stimuli can be found in appendix B.

- (7) a. Each new adventure is an experience that at the start will feel uncomfortable.  
b. Each new adventure is an experience that will feel uncomfortable at the start.
- (8) a. That gray squirrel was a creature that would bury nuts near the tree.  
b. ?That gray squirrel was a creature that near the tree would bury nuts.

### 3.2.3 Comprehension and attention checks

One instruction comprehension check question was included to indicate whether participants had read the instructions. It asked whether the task involved rating sentences on a scale, reading sentences and answering content questions, attempting to replace a missing word in sentences, or comparing two sentences (the correct answer). Two attention check questions were also included in the course of the survey to ensure that participants were not inattentively clicking through the survey. These checks asked participants to choose a specific response. The full text of the instruction comprehension check and attention checks can be found in appendix C.

### 3.3 Procedure

After clicking the survey link on the Amazon Mechanical Turk task page, participants were brought to the Ibex Farm website, where they accepted a consent form and read a set of instructions according to the condition to which they were assigned. After this, they answered the instruction comprehension question and began the experiment. On each trial, participants had to choose one of two sentences according to the instructions previously presented, which they could access at any time by clicking a link on the corner of the screen. After validating their choice, the next trial was started immediately until all of the critical sentence stimuli pairs, fillers, and both attention checks had been administered. Participants progress in the study was halted if they failed the instruction comprehension check or an attention check question in order to prevent having to reject these workers later, though this was not implemented until the second round of data collection. Following this, a brief demographic survey was administered addressing age, gender, native language, level of education, level of education of the most educated parent, country of origin, and whether the participant had ever taken a linguistics course. When this was completed, a randomly-generated code was presented along with the debriefing form. This code was copied and pasted into an entry field on the Amazon Mechanical Turk task page so that participants could receive compensation for their participation.

### 3.4 Analysis

Mixed-effects regression models are a powerful statistical tool for grammaticality judgments as they can accommodate random effects, such as participant, into one statistical model; as this was a forced-choice task which yielded binary data (whether or not a participant responds according to a prescriptive rule), a logistic regression was the appropriate mixed-effects model (Myers 2009). In this mixed-effects logistic regression, participant was set as a random factor and instruction condition (INTUITIVE SCHOOL, PROFESSOR, or TUTOR coded as three separate binary variables) were set as fixed factors. The baseline condition (i.e., the reference condition against which other conditions were compared) was thus the INTUITIVE PLAIN condition. The analysis was conducted using the `glmer` function of the `lme4` package (Bates et al. 2015) for the statistical software R (R Core Team 2013).

## 4 Results

### 4.1 Prepositional Phrasal Verbs

The hypothesis for the prepositional phrasal verb (PPV) sentence pairs was that participants would be more likely to choose a prescriptive response in the PROFESSOR condition. The full results of my logistic regression model on the PPV data can be found in table 1. Consistent with this hypothesis, it was found that the log odds of a prescriptive response were higher in the PROFESSOR condition than in the baseline condition,  $\beta = 0.6778$ ,  $z = 2.208$ ,  $p < .05$ .

### 4.2 Particle-Prepositional Phrasal Verbs

The hypothesis for the particle-prepositional phrasal verb (PPPV) sentence pairs was also that participants would be more likely to choose a prescriptive response in the PROFESSOR condition. As shown in table 2, the results of the logistic regression were not consistent with this hypothesis. It was found, however, that the log odds of a prescriptive response were higher in the TUTOR condition than in the baseline condition,  $\beta = 1.4082$ ,  $z = 2.448$ ,  $p < .05$ .

Predictor	$\beta$	Standard Error $\beta$	$z$ Value	$p$
Baseline (Intuitive Plain)	-1.0803	0.2172	-4.973	6.59e-07***
Intuitive School	-0.1457	0.3094	-0.471	0.6377
Professor	0.6778	0.3069	2.208	0.0272*
Tutor	-0.3445	0.3100	-1.111	0.2665

*Note:* \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$

Table 1: Logistic regression analysis of prepositional phrasal verb data.

Predictor	$\beta$	Standard Error $\beta$	$z$ Value	$p$
Baseline (Intuitive Plain)	-4.8859	0.4785	-10.211	< 2e-16 ***
Intuitive School	0.4316	0.5874	0.735	0.4625
Professor	1.0068	0.5812	1.732	0.0832
Tutor	1.4082	0.5752	2.448	0.0143 *

*Note:* \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$

Table 2: Logistic regression analysis of particle-prepositional phrasal verb data.

## 5 Discussion

The aim of this study was to reproduce the study described in Cowart (1997), which suggested that instructions (prescriptive PROFESSOR vs. INTUITIVE) may not have an effect on participant responses in grammaticality judgment tasks. To address the limitations of that study, the current study used approximately five times the number of participants, the more statistically powerful forced-choice task, and stimuli that directly compare sentences that follow a prescriptive rule with sentences that do not. Prepositional phrasal verb pairs, which can grammatically feature a raised or word-final preposition, were created to investigate whether participants were more likely to make a prescriptive response when both sentences were grammatical. It was hypothesized that, due to associations between prescriptivism and English professors, that in the PROFESSOR condition participants would be more likely to choose a prescriptive response, and this hypothesis was confirmed by the data, indicating that the prescriptive instructions do influence participants' responses.

Particle-prepositional phrasal verb pairs, the particle and preposition of which cannot both be raised grammatically, were also constructed; in this case, a prescriptive response would mean the participant had chosen the ungrammatical sentence, thus over-applying the prescriptive rule. It was again hypothesized that participants in the PROFESSOR condition would be more likely to choose the (ungrammatical) prescriptive response. Here, however, a significant difference was not found between the odds of a prescriptive response in the PROFESSOR condition and the baseline condition. It was found, however, that although the TUTOR instruction condition was not a significant predictor of prescriptive responses for the PPV sentence pairs, for the PPPV pairs, participants in the TUTOR condition were significantly more likely to respond prescriptively. Due to the exploratory nature of this condition, however, future research will be needed to confirm this effect, but it may suggest that a TUTOR condition is perhaps not a middle ground between intuitive and prescriptive instructions.

Finally, the study also attempted to investigate whether mentioning "school grammar" in IN-

TUITIVE SCHOOL instructions made a difference for participants' performance, and no such effect was found, demonstrating that the mention of "school grammar" likely does not prime participants to consider it despite explicit instructions to ignore it. Conversely, it may suggest that explicitly telling participants to disregard school grammar may not be necessary when asking for a "gut reaction."

### **5.1 Reevaluating the role of participant instructions**

Cowart conceded that while his null finding did reflect the general attitude of syntacticians that instructions do not matter, it remains an open empirical question. He outlines two principles for future research to demonstrate the importance of instructions:

First, they will need to show that giving informants instructions other than those that ask for a global, intuitive assessment of the sentence produces consistent, systematic differences of pattern in informant responses. Second, they will need to diagnose those differences that arise. That is, to make those differences of pattern fully useful, investigators will need to demonstrate that one sort of instruction makes informants comparatively more (or less) sensitive to specific aspects of the presented sentences (e.g., to processing difficulty associated with, say, rare vocabulary or known parsing complexity, or to semantic factors connected with reference or the empirical truth of the sentence). (Cowart 1997: 58–9)

For the PPV pairs, the present study has shown a systematic difference between participants in the PROFESSOR condition and those in the baseline condition. The PPPV sentence pairs were designed to address the second part, whether the instruction condition could cause participants to prioritize prescriptive correctness over grammaticality. For participants in the TUTOR condition, this does appear to be the case. Although this finding does not match the intuitions of syntacticians about the role of instructions, Schütze (2016) presented evidence that other task-related factors, including order, repetition, register, and speed of judgment can influence this task; in light of these findings, it would be rather surprising if instructions did not have an effect on participant judgments.

### **5.2 Limitations and directions for future research**

One potential criticism is that perhaps the design of the present study is too different from that of Cowart's to be directly comparable. Cowart's study, for example, used a Likert-type scale rather than a forced choice between sentence pairs. I argue that a forced-choice task is more appropriate, especially for exploratory research in this area, as it is more statistically powerful, allows for the direct manipulation of the presence of a prescriptive rule, and minimizes differences between sentence pairs, such as overall word frequency, slight differences in syntactic structure, sentence length, and the base grammaticality of sentences. I predict that the pattern found here would hold in a Likert-type rating task, but again, because a forced-choice task is more statistically powerful, it may require a larger number of participants and stimuli than are typically used in small-scale judgment studies, though the rise of web-based methods (such as those presented in this study) make cheaply and quickly recruiting large numbers of participants increasingly more feasible. Crucially, however, the stimuli in such a Likert-type rating task must be carefully controlled for the factors mentioned above in order to prevent confounds when sentences are presented individually.

Similarly, one might argue that Cowart (1997) did not directly manipulate the prescriptiveness of sentences in his study because, as many judgment tasks are not attempting to investigate phenomena that are confounded with prescriptive preferences, it is simply too contrived a situation. The examples from Schmidt and McCreary (1977), especially (2), illustrate why that is not always the case. For instance, even if a variation of the sentence in (2a) were being presented in order to investigate another syntactic phenomenon, a non-obvious extralinguistic factor, such as the presence of *There is* instead of *There are*, may lead to lower ratings for that sentence. This is one situation where creating instructions aimed at reducing the likelihood of a prescriptive response could lead to better experimental control. Another area where prescriptivist responses are particularly important is in the study of nonstandard dialects and minority languages, as Henry (2005) suggested. In the relatively formal setting of an academic study, mitigating the effect of prescriptivist attitudes that may arise from the more prestigious language or dialect would help ensure that participants are giving their intuitive judgments rather than judgments based on what they believe is “correct.” These findings suggest that instructions, at least avoiding prescriptive instructions, may be one way to achieve this.

While it can generally be assumed that the participants on Amazon Mechanical Turk represent a more diverse sample than the children and undergraduate students who typically participate in laboratory-based grammaticality judgment studies, it remains an open question to what extent these results are generalizable to other populations. As mentioned above, a key population in which to study the role of instructions will be speakers of nonstandard dialects or minority languages, because, as the findings of Henry (2005) indicate, mitigating prescriptive responses in this context could be significant for the reliability of results.

Additionally, the preference for prescriptive responses in the PROFESSOR condition for prepositional phrasal verbs suggests that participants are not only aware of prescriptive rules, but understand the role of education in distributing and enforcing these rules. One may suspect that level of education plays a role in this understanding, though it is possible to imagine this unfolding in two ways: Schmidt and McCreary (1977) suggested that perhaps because more educated participants have more exposure to prescriptive rules, this may make them more likely to respond prescriptively. On the other hand, perhaps more educated participants have more exposure to the idea that prescriptive rules are merely arbitrary concepts unrelated to speaker intelligence and thus feel less pressured to respond prescriptively overall. A large-scale study with groups from various education levels could help to disentangle these possibilities and suggest appropriate ways to frame instructions when working with populations that differ in level of education.

Finally, it might be expected that age plays a role in whether participants are more likely to respond prescriptively, as the emphasis on prescriptivism in the educational system may change over time. A study similar to the level of education study with various age categories could establish whether the effect of prescriptive instructions may interact with age.

## 6 Conclusion

Previous research on the grammaticality judgment task has suggested that task-related experimental control is important for securing stable and reliable grammaticality judgments. This study investigated whether participants who received prescriptive instructions were more likely to respond prescriptively than participants who were asked to give an intuitive response or respond as a peer English tutor. It was found that when both sentence choices were grammatical, participants in the professor condition were more likely to choose the prescriptive response. However, when the

response that was ostensibly prescriptively correct was ungrammatical, participants in the TUTOR instruction condition were more likely to select that ungrammatical answer. These results suggest not only that prescriptive instruction should be avoided when collecting grammaticality judgments, but that a TUTOR condition may not be a satisfactory alternative to those prescriptive instructions. Future research should focus on the effect of instruction type with specific populations, including speakers of nonstandard dialects or minority languages and speakers of varying age and level of education. This line of research will contribute to ensuring that the source of grammaticality judgments on which much of descriptive and theoretical linguistic work is based is empirically grounded and reliable.

## 7 References

- Bader, M. and Haeussler, J. (2010). Toward a model of grammaticality judgments. *Journal of Linguistics* 46, 273–330.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1–48.
- Bradac, J.J., Martin, L.W., Elliott, N.D., and Tardy, C.H. (1980). On the neglected side of linguistic science: Multivariate studies of sentence judgment. *Linguistics* 18, 967–995.
- Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. London: Sage.
- Curzan, A. (2014). *Fixing English: Prescriptivism and language history*. Cambridge: Cambridge University Press.
- Dąbrowska, E. (2010). Naive v. expert intuitions: An empirical study of acceptability judgments. *The Linguistic Review* 27, 1–23.
- Drummond, A. (2013). Ixex Farm. Online server: <http://spellout.net/ibexfarm>.
- Heck, F. (2008). *On pied-piping: Wh-movement and beyond*, *Studies in generative grammar*, volume 98. Berlin: Walter de Gruyter.
- Henry, A. (2005). Non-standard dialects and linguistic data. *Lingua* 115, 1599–1617.
- Labov, W. (1975). Empirical foundations of linguistic theory. In R. Austerlitz (ed), *The Scope of American Linguistics*, 77–133. Ghent: Peter De Ridder.
- Love, N. (1989). Language and the science of the impossible. *Language & Communication* 9, 269–287.
- Myers, J. (2009). The design and analysis of small-scale syntactic judgment experiments. *Lingua* 119, 425–444.
- Peer, E., Vosgerau, J., and Acquisti, A. (2013). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods* 46, 1023–1031.
- R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Riemer, N. (2009). Grammaticality as evidence and as prediction in a Galilean linguistics. *Language Sciences* 31, 612–633.
- Ross, J.R. (1979). Where’s English? In C.J. Fillmore, D. Kempler, and W.S.Y. Wang (eds), *Individual Differences in Language Ability and Language Behavior*, 127–163. New York: Academic Press.
- Schmidt, R.W. and McCreary, C.F. (1977). Standard and super-standard English: Recognition and use of prescriptive rules by native and non-native speakers. *TESOL Quarterly* 11, 415–429.
- Schnoebelen, T. and Kuperman, V. (2010). Using Amazon Mechanical Turk for linguistic research.

- Psihologija* 43, 441–464.
- Schütze, C.T. (2016). *The empirical base of linguistics: grammaticality judgments and linguistic methodology*. Berlin: Language Science Press.
- Schütze, C.T. and Sprouse, J. (2014). Judgment data. In R.J. Podesva and D. Sharma (eds), *Research methods in linguistics*, 27–50. Cambridge: Cambridge University Press.
- Smith, P.A. (2011). Attention, working memory, and grammaticality judgment in typical young adults. *Journal of Speech, Language & Hearing Research* 54, 918–931.
- Snyder, W. (2000). An experimental investigation of syntactic satiation effects. *Linguistic Inquiry* 31, 575–582.
- Spencer, N.J. (1973). Differences between linguists and nonlinguists in intuitions of grammaticality-acceptability. *Journal of Psycholinguistic Research* 2, 83–98.
- Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43, 155–167.
- Valian, V. (1982). Psycholinguistic experiment and linguistic intuition. In T. Simon and R. Scholes (eds), *Language, mind, and brain*, 179–188. Hillsdale, NJ: Lawrence Erlbaum.
- Zehr, J. and Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX).

## **Appendix A: Task instructions**

### **A.1 Intuitive Plain Condition**

Each page of this survey will present you with two sentences. Please read each of the sentences carefully.

For each sentence, you will indicate your gut reaction. If you think a sentence seems fully normal and understandable to you, select that sentence by clicking on it. In other words, if you think it is very odd, awkward, or difficult for you to understand, select the other sentence. When you are satisfied with your selection, click “Validate” to move on to the next page.

Note that you will only be able to select one of the two sentences.

### **A.2 Intuitive School Condition**

Each page of this survey will present you with two sentences. Please read each of the sentences carefully.

For each sentence, you will indicate your gut reaction. If you think a sentence seems fully normal and understandable to you, select that sentence by clicking on it. In other words, if you think it is very odd, awkward, or difficult for you to understand, select the other sentence. When you are satisfied with your selection, click Validate to move on to the next page.

Note that you will only be able to select one of the two sentences.

THERE ARE NO “RIGHT” OR “WRONG” ANSWERS. Please base your responses solely on your gut reaction, not on rules you may have learned about what is “proper” or “correct” English.

### **A.3 Professor Condition**

Each page of this survey will present you with two sentences. Please read each of the sentences carefully.

For each sentence, you will indicate whether or not you think the sentence is a well-formed, grammatical sentence of English. Suppose one of these sentences were included in a term paper submitted for a senior-level college English course that is taken only by English majors: would you expect the professor to accept this sentence?

If you think a sentence would be accepted, select that sentence by clicking on it. In other words, if you think a sentence would not be regarded as grammatical English by an appropriately trained person, select the other sentence. When you are satisfied with your selection, click “Validate” to move on to the next page.

Note that you will only be able to select one of the two sentences.

### **A.4 Peer Tutor Condition**

Each page of this survey will present you with two sentences. Please read each of the sentences carefully.

For each sentence listed, you will indicate whether or not you think the sentence seems English-sounding. Suppose that you are tutoring a friend who speaks English as a second language: would you tell them that this sentence makes them sound like a native speaker?

If you think that a sentence sounds native-like, select that sentence by clicking on it. In other words, if you think that a sentence sounds strange or unnatural, select the other sentence. When you are satisfied with your selection, click “Validate” to move on to the next page.

Note that you will only be able to select one of the two sentences.



## Appendix B: Sentence stimuli

### B.1 Prepositional phrasal verb sentence pairs

- (1) a. The brilliant compromise is a solution which both sides could happily agree on.  
b. The brilliant compromise is a solution on which both sides could happily agree.
- (2) a. The exhausting trip was a vacation which my friends will barely recover from.  
b. The exhausting trip was a vacation from which my friends will barely recover.
- (3) a. The antique painting was a piece which the critic would closely stare at.  
b. The antique painting was a piece at which the critic would closely stare.
- (4) a. The careful argument was a statement which no opponent could easily argue with.  
b. The careful argument was a statement with which no opponent could easily argue.
- (5) a. The charming neighborhood was a district which young couples could comfortably live in.  
b. The charming neighborhood was a district in which young couples could comfortably live.
- (6) a. The new plan is a project which the mayor should immediately work on.  
b. The new plan is a project on which the mayor should immediately work.
- (7) a. The monthly journal is a magazine which all experts should definitely subscribe to.  
b. The monthly journal is a magazine to which all experts should definitely subscribe.
- (8) a. The rare bird is a species which every biologist should enthusiastically teach about.  
b. The rare bird is a species about which every biologist should enthusiastically teach.
- (9) a. The controversial issue was a topic which few activists would openly fight for.  
b. The controversial issue was a topic for which few activists would openly fight.
- (10) a. The new method is a technique which all surgeons should eagerly learn about.  
b. The new method is a technique about which all surgeons should eagerly learn.
- (11) a. Her impatient sister is a companion whom few people would willingly travel with.  
b. Her impatient sister is a companion with whom few people would willingly travel.
- (12) a. My intimidating professor is an expert whom no student can confidently speak to.  
b. My intimidating professor is an expert to whom no student can confidently speak.
- (13) a. Her well-behaved daughter is an angel whom most babysitters would gladly care for.  
b. Her well-behaved daughter is an angel for whom most babysitters would gladly care.
- (14) a. That retired teacher is the customer whom the server would often wait on.  
b. That retired teacher is the customer on whom the server would often wait.
- (15) a. The careful bicyclist is an athlete whom no accident could ever happen to.  
b. The careful bicyclist is an athlete to whom no accident could ever happen.
- (16) a. That foster parent was a guardian whom many children could reliably depend on.  
b. That foster parent was a guardian on whom many children could reliably depend.
- (17) a. The master violinist was a musician whom few professionals could really compare to.  
b. The master violinist was a musician to whom few professionals could really compare.
- (18) a. The friendly star was a figure whom many fans would regularly write to.  
b. The friendly star was a figure to whom many fans would regularly write.
- (19) a. The unpredictable politician was a candidate whom most people would never vote for.  
b. The unpredictable politician was a candidate for whom most people would never vote.
- (20) a. The offensive comedian was an entertainer whom no audience would generally laugh at.  
b. The offensive comedian was an entertainer at whom no audience would generally laugh.

**B.2 Particle-prepositional phrasal verb sentence pairs**

- (1) a. That rare edition was a book which the collector would always look out for.  
b. That rare edition was a book out for which the collector would always look.
- (2) a. This new technology is a subject which all professionals must tirelessly keep up with.  
b. This new technology is a subject up with which all professionals must tirelessly keep.
- (3) a. The first snowfall was an event which my brother would impatiently look forward to.  
b. The first snowfall was an event forward to which my brother would impatiently look.
- (4) a. The popular shampoo was a product which the store would regularly run out of.  
b. The popular shampoo was a product out of which the store would regularly run.
- (5) a. This secure cell was an area which no prisoner could ever break out of.  
b. This secure cell was an area out of which no prisoner could ever break.
- (6) a. Her first birthday is an event which her dad must unfortunately miss out on.  
b. Her first birthday is an event out on which her dad must unfortunately miss.
- (7) a. The unsigned contract is a deal which the businesswoman could still back out of.  
b. The unsigned contract is a deal out of which the businesswoman could still back.
- (8) a. The upcoming release is an update which the programmer will gladly help out with.  
b. The upcoming release is an update out with which the programmer will gladly help.
- (9) a. The difficult hike is a challenge which any adventurer should carefully gear up for.  
b. The difficult hike is a challenge up for which any adventurer should carefully gear.
- (10) a. The tough test was a milestone which no student could ever feel up to.  
b. The tough test was a milestone up to which no student could ever feel.
- (11) a. My friendly uncle is a relative whom most everyone can easily get along with.  
b. My friendly uncle is a relative along with whom most everyone can easily get.
- (12) a. This kind woman was a friend whom my aunt would frequently catch up with.  
b. This kind woman was a friend up with whom my aunt would frequently catch.
- (13) a. Their dishonest mother was a liar whom the children could never make up with.  
b. Their dishonest mother was a liar up with whom the children could never make.
- (14) a. A real friend is someone whom you must consistently stick up for.  
b. A real friend is someone up for whom you must consistently stick.
- (15) a. That timid accountant was the employee whom the boss would always talk down to.  
b. That timid accountant was the employee down to whom the boss would always talk.
- (16) a. That bearded man is the teacher whom the substitute would occasionally fill in for.  
b. That bearded man is the teacher in for whom the substitute would occasionally fill.
- (17) a. That tall carpenter was the gentleman whom my sister would sometimes go out with.  
b. That tall carpenter was the gentleman out with whom my sister would sometimes go.
- (18) a. The quiet child is a troublemaker whom all babysitters should often check in on.  
b. The quiet child is a troublemaker in on whom all babysitters should often check.
- (19) a. The expert chef was a master whom any amateur must humbly look up to.  
b. The expert chef was a master up to whom any amateur must humbly look.
- (20) a. Your strange cousin is a person whom the family must reluctantly put up with.  
b. Your strange cousin is a person up with whom the family must reluctantly put.

### B.3 Filler sentence pairs

- (1) a. The talented musician was a performer who on some days would sing opera.  
b. The talented musician was a performer who would sing opera on some days.
- (2) a. Each new adventure is an experience that at the start will feel uncomfortable.  
b. Each new adventure is an experience that will feel uncomfortable at the start.
- (3) a. That old song was a tune that by the end would sound familiar.  
b. That old song was a tune that would sound familiar by the end.
- (4) a. That abstract painting was a work that after a minute would look boring.  
b. That abstract painting was a work that would look boring after a minute.
- (5) a. The delicate pastry is a food that before too long could lose flavor.  
b. The delicate pastry is a food that could lose flavor before too long.
- (6) a. A guard dog is a pet that during the night must keep watch.  
b. A guard dog is a pet that must keep watch during the night.
- (7) a. That quiet teenager was an employee who at closing time would wash dishes.  
b. That quiet teenager was an employee who would wash dishes at closing time.
- (8) a. The beautiful wedding was an event that for some time will be remembered.  
b. The beautiful wedding was an event that will be remembered for some time.
- (9) a. That stubborn stain is an annoyance that by this point should be gone.  
b. That stubborn stain is an annoyance that should be gone by this point.
- (10) a. The chocolate cake is a dessert that later this evening must be baked.  
b. The chocolate cake is a dessert that must be baked later this evening.
- (11) a. The young man was a student who in the library would read books.  
b. The young man was a student who would read books in the library.
- (12) a. The older customer was a patron who at this cafe would sip coffee.  
b. The older customer was a patron who would sip coffee at this cafe.
- (13) a. The stern captain is the officer who on the boat could give orders.  
b. The stern captain is the officer who could give orders on the boat.
- (14) a. The young intellectual is a writer who in the park would write poetry.  
b. The young intellectual is a writer who would write poetry in the park.
- (15) a. The serious weightlifter was an athlete who at the gym would rarely speak.  
b. The serious weightlifter was an athlete who would rarely speak at the gym.
- (16) a. The new teacher is a vocalist who at this school will teach music.  
b. The new teacher is a vocalist who will teach music at this school.
- (17) a. The amateur chef was a cook who in the kitchen would develop recipes.  
b. The amateur chef was a cook who would develop recipes in the kitchen.
- (18) a. The experienced artist is a master who by the statue will paint portraits.  
b. The experienced artist is a master who will paint portraits by the statue.
- (19) a. That gray squirrel was a creature that near the tree would bury nuts.  
b. That gray squirrel was a creature that would bury nuts near the tree.
- (20) a. The post office was a service that around the world would deliver letters.  
b. The post office was a service that would deliver letters around the world.

## **Appendix C: Instruction comprehension and attention check questions**

### **C.1 Instruction comprehension question**

For this survey, it is very important that you understand the instructions. Please select the answer that best describes what you will be doing in this task and then click “Start the experiment!” If you would like to review before you answer, click on the “Instructions” link above.

- Attempting to find the missing word in each sentence.
- Reading sentences and then answering questions about their content.
- Rating sentences on a scale from 1 to 7.
- Selecting one of two sentences according to the instructions on the previous page.  
(Correct answer.)

### **C.2 Attention check questions**

- (1) a. If you’re reading this, please select this sentence  
b. If you’re reading this, please select the other sentence.
- (2) a. If you’re reading this, do not select the other sentence, select this one.  
b. If you’re reading this, do not select this sentence.