USING DEVELOPMENTAL SEQUENCES TO ESTIMATE ABILITY WITH ENGLISH GRAMMAR: PRELIMINARY DESIGN AND INVESTIGATION OF A WEB-BASED TEST[1]

JOHN M. NORRIS

University of Hawai'i at Manoa

The usefulness of open-access, web-based language learning resources can be enhanced considerably by aligning learner skill or proficiency levels with materials, activities, lessons, and other instructional media that are gauged accordingly. This paper reports on the design of an English grammar test for the purpose of making broad initial distinctions among learner ability levels, such that subsequent web-based assessment and learning tools could be articulated to their needs. In order to estimate initial learner differences within a very brief period of testing time, and in a format amenable to the online environment, developmental sequences in the acquisition of English morphosyntax were investigated as a basis for test design. Resulting test items tapped learners' abilities to apply five word order rules in a variety of linguistic contexts; these items were piloted with 57 English learners, from pre-academic to university-entrance levels of proficiency. Findings indicated that the developmentally most advanced items (testing (a) the cancel-inversion rule in embedded questions and (b) relative clauses) discriminated well among the different proficiency levels in the population sample; in addition, relatively reliable twelve-item tests could be constructed on the basis of the items investigated. In conclusion, recommendations for follow-up web-based test implementation and validity evaluation are provided. Pilot test items and forms are also appended.

INTRODUCTION

This paper reports on the development of one sub-test in a battery of computerized tests intended to direct diverse English language learners to web-based or web-accessible materials and activities appropriate to their interests and general L2 ability levels. Dramatic increase in the availability—not to mention variety, quality, and L2 focus—of

[1] This research was supported by Pearson Education. My thanks to Carol Chapelle and Joan Jamieson for inviting me to contribute this grammar sub-test to the "Test Your English" project, the outcomes of which have been published recently under the name *Longman English Assessment*. Many thanks also to Steve Jacques, who contributed not only substantial time and effort to data collection but also critical insights into practical test design concerns. Finally, I am indebted to Lourdes Ortega, who suggested a number of useful items types for testing learners' word order competences.

Second Language Studies, 24(1), Fall 2005, pp. 24-128.

web-based language learning resources has generated considerable current demand for assessments which can quickly match learners with relevant tools, activities, and other components of self-access L2 instruction (e.g., Ingold, 2002). Though good work has been done in defining the parameters and possibilities of web-based language assessment (e.g., Roever, 2001), the online medium certainly limits what aspects of L2 knowledge and ability may be assessed in what ways (e.g., Norris, 2001), and it remains to be seen how specific assessment purposes will best be met.

As detailed in Chapelle, Jamieson, and Hegelheimer (2003), where learners are seeking access to well-defined and interrelated language learning materials, assessment can be used to carefully articulate learners' abilities and interests with those reflected in available materials. They describe how, in conjunction with a publisher's extensive online English language learning offerings, a sequence of assessments can be developed in order to separate learners into increasingly narrow ability and interest groups. Once so identified, these learners are then directed to the corresponding types of available and appropriate materials on the publishers' web site. The current paper reports on the initial design and investigation of one component in the overall assessment battery that they outline.

The first section of this assessment battery comprises an *Ability Level Finder*, which utilizes several estimates of global language ability—specifically, two brief tests of vocabulary and grammar knowledge—in order to make initial, automated distinctions among a range of learners who would potentially seek access to the publishers' web-based resources. These quick initial distinctions are intended to link examinees with the ability-appropriate portions of subsequent, more detailed skills-based tests. Accordingly, the project reported here took as its goal the development of a practical and efficient ESL/EFL grammar test instrument which could be implemented on-line and used to distinguish among language learners across a very broad spectrum of ability levels in a short amount of testing time.

In this report, I first provide a brief rationale for the pilot test items and instruments developed in the project. In the following section I describe the methodology followed in developing items and investigating their qualities, including the analyses of pilot test outcomes. Subsequently, I detail possibilities for improving items and tests through specific revisions, and I offer two proposals for the development of three, short but still reliable, 12-item operational test forms. Finally, I suggest several areas in need of further research and development efforts. Also appended to this report are all components of the pilot test instrument, three complete revised test forms intended as possible operational tests for subsequent use, and related documents used in the study.

Rationale for Test and Item Development

Given the intended uses of the *Ability Level Finder* described above, the first task in the current project was to identify a rationale on which to base test design and item selection for a reliable and useful grammar test. At the same time, constraints associated with assessment purpose and context had to be kept in mind; namely, test administration, scoring, and decisions needed to occur as efficiently as possible within a web-based test delivery environment. Therefore, the question addressed at this stage in test development asked what features of the (English) grammatical system would lend themselves to informing the kinds of broad distinctions required, while doing so on the basis of only a few computer-scored items and in a very short amount of time. Selection of a particular grammatical subsystem of English (e.g., morphological marking of verbs for person, number, and time) did not present a particularly promising strategy, given the variability with which such rule-governed behavior may be learned/acquired by learners under differing instructional and/or naturalistic circumstances, as well as the ambiguous relationship between a single subsystem of this sort and the more global kinds of language ability interpretations required for the current test. For similar reasons, any unmotivated constellation of a few items each for a variety of grammar rules (such as those found in many commercial English language textbooks) would lead most likely to high variability, and therefore unpredictability, in performance by the range of actual learners completing the test. Alternatively, covering substantial portions of the L2 grammar system would require a very large number of test items and too much administration time. On what basis, then, could a finite set of English grammar items be selected, such that they would produce predictable and reliable variance among examinees that could be associated closely with differences in estimates of global language ability?

One potentially relevant area of theory-motivated inquiry into L2 grammatical development has focused on learners' acquisition of word order rules in several languages. Initial cross-sectional and longitudinal empirical research associated with the ZISA project (see Clahsen, Meisel, & Pienemann, 1983; Meisel, Clahsen, & Pienemann, 1981) identified an apparent implicational relationship in the emergence of L2 learners' abilities to use (productively) several inflexible rules of German word order. Subsequent work (e.g., Clahsen, 1984; Pienemann, 1998; Pienemann & Johnston, 1987) proposed an expanded justification, grounded in theories of cognitive processing and lexical-functional grammar, which accounts for the invariant order of emergence for these and similar syntactic rules in a variety of second languages, including English (see Johnston, 1985; Pienemann, Johnston, & Brindley, 1988; Pienemann & Mackey, 1993). According to these theories, and supported by accumulated empirical findings over some 20 years of research (see review in Pienemann, 1998), regardless of L1 or instructional/acquisitional context differences, all learners of a given L2 produce initial instances of particular word order rules according to fixed implicational sequences.

In its most recent and thorough formulation within processability theory (Pienemann, 1998), the following implicational hierarchy is proposed for the emergence of syntactic phenomena in L2 English development: (a) single constituent or word-level syntax (i.e., no word order); (b) canonical (Subject Verb Object) word order; (c) adverb-fronting, topicalization, and pre-verbal negation; (d) subject-verb inversion in yes/no questions; (e) movement of auxiliary verbs to the position following a WH question word; and (f) the reversal of subject-verb inversion in indirect questions. The theory also makes similar proposals for the emergence of several morphological rules which are linked to word order stages on the basis of common underlying cognitive processing procedures and constraints. Based on these theoretical predictions, and associated empirical support, many have recommended that developmental sequences be used to inform the teaching of second and foreign languages, in syllabus construction, and for instruction-related assessment purposes (e.g., Long & Crookes, 1993; Pienemann, 1985, 1989; Pienemann & Johnston, 1987; Pienemann, Johnston, & Brindley, 1988), and some empirical research has indicated their relevance in instructional settings (e.g., Ellis, 1989; Kanagy, 1994; Pienemann, 1987; Spada & Lightbown, 1999).

Despite a solid theoretical foundation and empirical evidence in support of claims about the ordered emergence of L2 word order phenomena, some have expressed doubts about the extent to which such developmental sequences may play a practical role in language teaching and assessment (e.g., Hudson, 1993; Mellow, 1996; Norris, 1996). In language assessment, there are several major impediments to applying developmental sequences for practical uses. First, while the emergence of word order patterns occurs in an invariant sequence, with one rule always preceding the next, the linguistic and communicative contexts in which such rules will first be used may vary substantially from learner to learner. Thus, the elicitation of very initial stages of productive use of a given rule may be quite unpredictable on a learner by learner basis, and often the only means for achieving a warranted interpretation about a learner's developmental level according to these stages is through the collection and careful analysis of very large amounts of free production data (a requirement which does not fare well in light of the constraints posed by many language assessment contexts, although see interesting methods for streamlining this process in Pienemann, Johnston, & Brindley, 1988; Pienemann & Mackey, 1993).

A second problem for language assessment is that measuring the emergence of a rule, a cognitive construct, requires some predetermined behavioral criterion which can be observed and interpreted with consistency (see discussion in Norris & Ortega, 2003). The link between the two has proved controversial for inquiry into developmental sequences, with various researchers proposing differing sets of criteria for what counts as evidence for initial productive use (emergence) of a given rule (see, e.g., discrepancies between Ellis, 1989, and Pienemann, 1998). At the same time, developmental theory does not make predictions about the rates or routes with which productive *accuracy* (often referred to as mastery) of a given rule may be developed; thus, a theory-derived measure of word order development would currently be challenged to establish and justify behavioral criteria for determining emergence.

Finally, there is a related and key third problem in applying developmental sequences for meeting some language assessment purposes. Available research evidence indicates that *emergence* of the full range of posited developmental stages occurs very early among instructed language learners, though not necessarily among naturalistic learners, while the development of target levels of *accuracy* with these rules may carry on for some time (Ellis, 1989; Hudson, 1993; Norris, 1996). Developmental sequence measures which operationalize emergence-based scoring (i.e., based on initial productive use in any linguistic/communicative context) may only prove relevant for assessing a restricted range of beginning-proficiency (instructed) language learners. At this point, it remains unclear to what extent, if any, there exists a systematic and predictable relationship between developmental *emergence* of morphosyntactic rules and more global notions of language ability, proficiency, or communicative competence.

Although these observations may delimit the usefulness of developmental measures for meeting many assessment demands within language education contexts, some research has suggested that relationships do exist between levels of *accuracy* with developmental word order phenomena and broad global proficiency differences among learners. For example, Norris (1996) found that learners' target-like accuracy in producing developmentally sequenced German word order rules increased consistently in association with learners' oral proficiency levels (according to ratings on the ACTFL, 1986, *Proficiency Guidelines*). Furthermore, he found that word order rules at higher developmental stages were produced with high degrees of accuracy only among learners rated at the more advanced levels of oral proficiency. He also found that learners rated at *Intermediate-Low* and lower levels of oral proficiency did not produce any evidence of the highest predicted stage in German word order development (i.e., the movement of finite verbs to the end of a subordinate clause). In light of such findings, it may be the case that a focus on accuracy in the use of word order rules can provide an effective basis for distinguishing among broadly differing groups of L2 learners. Especially accuracy with rules at the upper stages in developmental sequences may serve as a useful basis for predicting broad proficiency differences, since such accuracy may indicate not only that examinees are able to process language at the multi-constituent sentential and inter-clausal level (and all of the preceding stages which are theoretically implied) but also that such processing is occurring efficiently enough for attention to grammatical accuracy to take place (see further discussion in Skehan, 1998).

For the purposes of the grammar section on the *Ability Level Finder*, items were developed to tap examinees' abilities in producing rules at stages two through six in the

L2 English word order developmental sequence listed above, and a target-like accuracy criterion was adopted for scoring item performances (see Methodology section). In addition, in order to address the possibility that items representing these stages might not prove sufficient for distinguishing among very advanced proficiency examinees, an additional set of items was derived by extending the processing rationale provided for the highest developmental stage in word order rules. Thus, given that processability theory (Pienemann, 1998) posits the inter-clausal exchange of grammatical information as the top level within L2 word order development, cognitive processing procedures associated with any production of relative clauses are most likely first engaged at this highest level within the developmental sequence. Therefore, items based on the word order required by different types of relative clauses might tap even higher degrees of processing, and therefore help to distinguish among more advanced examinees. Several types of relative clauses were incorporated as an additional basis for item development.

Finally, because test items were to be self-accessed by examinees, administered on-line, and scored automatically, free constructed-response item formats—the formats used in measuring productive developmental phenomena to date—were obviously out of the question. However, because word order phenomena were to serve as the constructs of interest for the test, an item format needed to be developed which enabled examinees to demonstrate knowledge (or the lack thereof) of the rules constraining movements of particular grammatical constituents. A constrained constructed-response format was chosen for item construction, to enable the focused and controlled elicitation of particular aspects of sentence-level language processing ability. Further, unlike selected-response item formats, a constructed-response item would not supply examinees with models of the accurate word order rules in the form of response options, thereby reducing the likelihood of learning or ‘catching on’ to the construct during the test (see related discussion in Norris & Ortega, 2003). In light of its compatibility with the eventual computer-mediated test delivery system, a “drag-and-drop” item format was selected, wherein grammatical constituents of a sentence could be presented individually in boxes and these boxes could be selected and placed by examinees into corresponding positions within a series of blanks constituting the target sentence (see description in Methodology section). This item format has been used successfully, in a paper-and-pencil mediated test, for eliciting word order phenomena in previous research (e.g., Spada & Lightbown, 1999).

METHOD

With the preceding considerations in mind, items were developed and a pilot test instrument compiled in order to initiate investigations into the extent to which the grammar component of the *Ability Level Finder* would produce interpretable and useful outcomes. This section describes pilot test development and research methods, including: (a) participants; (b) test materials; (c) test administration and scoring; and (d) data analyses.

Participants

Pilot-test volunteers were recruited from two separate English language service units at the University of Hawai‘i at Manoa (UH): the Hawaii English Language Program (HELP) and the English Language Institute (ELI), both operated by the Department of Second Language Studies. Prior to participant recruitment, permission was solicited from the directors of the two units, who reviewed and agreed to the proposed project and the activities to be undertaken by participants.

Students enrolled in HELP and ELI courses represent a broad range of English language proficiency levels. HELP provides general-purpose English language courses to a population of adult international students; these students are not otherwise matriculated in university courses and they range from novice learners to those preparing for university admissions requirements. HELP courses are divided into five ability levels (designated from beginner to advanced as level-10, -20, -30, -40, and -50 classes), and students are placed into classes at a given level on the basis of placement test scores or advancement from previous levels. Standardized proficiency test scores are not consistently available for HELP students; however, program personnel have designed the courses for abilities ranging from low-functional to pre-university needs.

The ELI serves a more advanced level of English language user (students must score between 500 and 600 on the TOEFL to be admitted), although teachers in both units report considerable overlap between HELP level-50 learners and ELI learners in terms of students' global English abilities. The ELI provides focused English for Academic Purposes courses to undergraduate and graduate international students; these students are also matriculated in other courses of study at UH. ELI students are granted admission to the university contingent upon the completion of a number of classes within the ELI (this number depends on the results of a placement exam). ELI courses are divided into two ability levels (designated as level-70 or -80 classes, with an additional undergraduate writing course carrying a different numeric designation, ELI 100, but also categorized in terms of ability as level 80).

With the objective of recruiting 50 pilot participants to represent much of the ability range of English language learners who might eventually take the "Test Your English" exam, volunteers were solicited from all levels of the HELP and ELI programs. In order to maximize the potential differences within such a broad target population, ten participants were sought from each of the following level groupings: (a) HELP level-10 and -20 students; (b) HELP level-30 and -40 students; (c) HELP level-50 students; (d) ELI level-70 students; and (e) ELI level-80 students. In response to recruitment efforts, fifty-seven students volunteered for pilot testing (and all were allowed to participate). Table 1 shows that recruitment expectations for each of the five learner levels were essentially fulfilled, with a number of additional learners volunteering for level-group 3. Each participant was compensated $10.00.

Table 1

Pilot Test Participants by Language Program Level Groupings

Participant level	*N*
(1) HELP 10/20	10
(2) HELP 30/40	19
(3) HELP 50	10
(4) ELI 70	9
(5) ELI 80	9
Total	57

In order to clarify the extent to which results of this pilot test investigation could serve as the basis for generalizations about the performance of likely future test takers, basic demographic data were solicited from participants (see Appendix A, *Background Information* form). Nearly equal numbers of participants reported Japanese (*N*=26) and Korean (*N*=23) as their first language, and a handful of other first languages were also reported: Chinese (3), French (1), Russian (1), Tetun/Portuguese (1), Thai (1), and Turkish (1). Clearly, the overwhelming majority of participants represented only two first language groups; as such, findings based on this pilot test may not generalize in their entirety to students with other first languages. Table 2 shows that participants ranged in age from 18 to 40 years, and the mean age of 25 years is reflective of the university student population (including both undergraduate and graduate students) solicited for the pilot test. Participants also reported widely varying amounts of previous English study, although, given the potentially radical differences in types of language instruction experienced by students, this variable probably should not be interpreted as showing more than the fact that virtually all students had experienced some prior English-language training. Nevertheless, because participants reported an average of six years of study, it may be the case that the average levels of language knowledge/proficiency of this pilot test sample exceed somewhat those of the target population of examinees for "Test Your English".

Table 2
Pilot Test Participant Characteristics

Statistic	Age (years)	English study (years)
Mean	25.05	6.42
SD	5.78	3.73
Min	18.00	0.00
Max	40.00	20.00

Finally, Table 3 shows average TOEFL scores where available. It should be noted that, while average TOEFL scores increase as participant level increases (and, based on program placement test scores, it can be assumed that the level-group 1 students would score on average lower than all other groups), there is also substantial variability evident within individual levels as well as considerable overlap in individual scores across all levels. Thus, while a wide range of proficiencies seems to be represented by the sample of 57 examinees who took the pilot test, it should not be assumed, on the basis of language program level status, that categorical English knowledge/proficiency differences existed between participants in the different level groups.

Table 3

Average TOEFL Scores by Participant Grouping

	Participant level-group				
Statistic	(1)	(2)	(3)	(4)	(5)
N (reporting)	--	9	8	9	8
Mean	--	455.22	511.50	520.44	534.13
SD	--	41.45	34.47	15.60	16.44
Min	--	418.00	470.00	503.00	520.00
Max	--	533.00	560.00	550.00	567.00

Test Materials

In order to operationalize and investigate the constructs in focus for the "Test Your English Grammar Ability Finder Pilot 1" (GP1), the following test materials were developed and compiled into a single paper-and-pencil test instrument: (a) participant consent form, (b) participant background information form, (c) general test instructions, and (d) seven test sections, each addressing different constructs and featuring unique item formats and instructions. Prior to describing each of these components of the pilot test, it should be noted that one overarching concern for pilot test development had to do with the recruitment of volunteer participants. Past experiences in recruiting volunteers from within the two English language service units at UH have shown that students are generally not willing to participate in research projects outside of class time, unless participation: (a) is compensated, (b) requires less than one hour of their time, and (c) occurs in easily accessible venues. Therefore, in addition to compensating volunteers and administering the pilot test at convenient times in the vicinity of students' classrooms, the test materials and test administration procedures were designed such that the process would likely require no more than one hour.

Consent, background information, general instructions. The entire pilot test instrument GP1 is reproduced in Appendix A. The first three pages show the title page, consent form, general instructions, and background information form. The *consent form* follows a standard format for soliciting volunteers' consent (and used in previous research at UH). It ensures the anonymity of participant data and the use of those data for research purposes only; all participants granted consent to the use of their test data for such purposes. *General test instructions* describe the test and administration procedures and were developed using basic language in order to facilitate understanding on the part of all participants. Instructions explain that examinees should work through each section and then wait before going on to the next section; they are also told not to go back and work on previous sections (and they are reminded to stop at the end of each section and wait for further instructions from the test administrator). These precautions were taken so that the test administrator could ensure that all examinees understood the differing requirements of the item formats in each test section. In addition, items in earlier sections of the test focused on constructs which could be revealed by items in subsequent sections (e.g., an item prompt in the form of a question would reveal word order rules for question formation that were being tested in a preceding section). Examinees were therefore not allowed to return to items within a section, once that section had been completed. Finally, instructions stressed that participants should feel free to ask any questions during the test in order to clarify what was required by the various item formats. A *background information form* also requested basic demographic data from participants, as reported in the preceding section on participants.

Test sub-sections. Seven unique test sections were developed with differing item formats and in order to tap specific constructs (see Appendix A). Each section and the corresponding formats and constructs are detailed below. Given the objective of identifying a subset of items which functioned similarly in distinguishing among broad groups of examinees, and in order to eventually sample three comparable 12-item operational test forms, numerous items were developed for each section. Business English content was used in roughly one-third of the items for each section in order to provide for one 12-item test form for a separate Business English subtest. Finally, it should be noted that example items were provided for each test section, in order to facilitate examinees' understanding of the performance required by a given item type. However, in order to avoid divulging the rule or construct itself to examinees, by modeling it within the example item (e.g., providing an example of a sentence with a relative clause for the section testing relative clauses), care was taken to develop examples which demonstrated the item format without revealing the particular rule(s).

Section 1. Items in this section were developed in order to provide a quick estimate of examinees' abilities with several morphological phenomena in English which are developmentally related to the word order rules in focus on Sections 2-6 of the pilot test. Thus, these items were included as a means for crosschecking performances on the items of interest for the operational test forms. The morphological rules addressed in Section 1 included: (a) plural marking on regular nouns following plural determiners (and the provision of plural forms for irregular nouns); and (b) plural/singular marking (and especially third person –s) on regular verbs following plural/singular nouns. Each item presented the examinee with a picture prompt (e.g., a boy playing soccer) followed by a short sentence with two blanks in the subject and finite verb positions. Examinees were instructed to write in a single word (spelled correctly) for each blank in order to complete a good sentence about the picture. Linguistic and pictorial contexts were provided in order to encourage examinees to produce the following: (a) four items required a singular noun and verb (provision of third person –s); (b) two items required a regular plural noun and verb; and (c) two items required an irregular plural noun and verb. An example item utilized a noun which does not mark the plural form (fish) and an irregular verb (to be), so as to reveal neither plural marking nor third person –s.

It was hypothesized that examinees who showed knowledge of both of the morphological rules tested would perform with high levels of accuracy on items in Sections 2 and 4 of the test, but with variable levels of accuracy on items in Sections 3, 5, and 6 (see below). Examinees who did not show evidence of the morphological rules were hypothesized to have less success with Sections 2 and 4, as well as great difficulty with the other test sections.

Section 2. Items in this section required students to combine words in order to “make a sentence” (as stated in the section title) using canonical English word order (CAN) and canonical word order with negation (CAN+N). Each item first presented examinees with a statement in quotation marks from a speaker who was given a descriptive title (e.g., classmate, friend, mother, boss, etc.). This statement was followed by a set of boxes, with each box containing one or several words (e.g., in the case of a prepositional phrase, the phrase was maintained in a single box); the boxes were arranged such that the target word order was scrambled. A set of blanks corresponding to the number of word boxes was presented in the following line, and this line was also given a speaker title (i.e., the person responding to the initial speaker’s statement). In addition, the blanks were marked with quotation marks and a period to indicate that they represented a statement by the labeled speaker. By way of example, item 10 is reproduced below.

Item 10:

Mother: “Sandra gets good grades in school.”

Words:	studies	a lot	she

Father: “___________ ___________ ___________.”

Examinees were instructed to rearrange the word boxes in order to make an appropriate response to the initial statement, and instructions stressed that words within one box should remain together in one blank in their answers (thereby providing a comparable response format to that which might be used in a computer-based test, wherein examinees would “drag and drop” word boxes into corresponding sentence blanks). Because it was impossible to provide an example item which did not model CAN (or subsequent test constructs), and owing to the probability that virtually all examinees would likely be able to construct basic subject-verb-object sentences, only three CAN items were included in this section (primarily as a means for introducing students to the item response format). An additional twelve items tested CAN+N (i.e., the suppliance of a negator following the finite verb) in a variety of linguistic contexts; both contracted negators (e.g., don't, won't, etc.) and analyzed negators (e.g., do not, no, etc.) were included.

Section 3. Items in this section asked examinees to "complete the sentence" that was started for them in item response stems. These nine items sought to test examinees' abilities to apply a cancel-inversion rule (X INV), which requires that subject-verb inversion in English questions be reversed in indirect questions (i.e., from VSO to SVO word order). Item formats differed in this section in that the response line now provided examinees with the first several words of a sentence (e.g., "I wonder...") followed by the appropriate number of blanks for the word boxes. The example item showed the "complete the sentence" response format with an indicative statement instead of an indirect question, in order to conceal the construct. By way of example, item 25 is reproduced below.

Item 25:

Travel agent: "The trip to Tokyo lasts seven hours."

Words: | a movie | will | they | show |

Traveler: "*I hope* ___________ ___________ ___________ ___________."

Section 4. Items in this section required examinees to "ask a question" in response to the item prompt statement by rearranging the word boxes. The format for these items was identical to items in Section 2, with the substitution of a large, bold-font question mark in the place of a period at the end of each response line. Items in Section 4 attempted to test two different question formation rules (with 13 items each): (a) subject-verb inversion in yes-no questions (Y-N INV); and (b) movement of finite auxiliary verbs in WH questions to the slot following the question word (Do-aux 2nd). Four additional experimental items were also developed in order to investigate whether potentially more difficult question formation rules would prove to distinguish better among examinees than Y-N INV and Do-aux 2nd questions. Two tag questions (i.e., a statement followed by a contracted affirmative-negative question, see items 58 and 61) and two affirmative-negative questions (i.e., questions which utilize an analyzed negator for emphatic purposes, see items 59 and 62) were included. The example item for this section modeled question formation with a Y-N INV type question (inverting “there is” to “is there”). While this example should not have proved prejudicial for Do-aux 2nd and subsequent items, it is possible that it did serve as a model for the Y-N INV questions. This issue is addressed further in the results section below. Item 33 is provided below as an example of the items in this section.

Item 33:

Girl: “I like to go to the beach.”

Words: volleyball | you | play | do

Boy: “___________ ___________ ___________ ___________ **?**”

Section 5. The nine items in this section utilized a format identical to that in Section 3, with the exception that the response stem now began with a question word and finished with a large question mark. Items in this section were intended to test the same cancel-inversion rule for indirect questions, but this time embedded within a second question (X INV+Q). The example item modeled the response format with a Y-N INV question, thereby concealing the construct of interest for the section.

Section 6. This final word order section utilized a unique item response format in order to tap examinees’ abilities to produce target-like relative clauses. For this section alone, the number of word boxes in the item prompt exceeded the number of response blanks by one. Thus, examinees were forced to select which of the boxes could combine together to form an accurate sentence. Six items tested two general types of relative clause formation in English: (a) pronoun-deletion clauses, wherein the relative pronoun can be left out of all except subject clauses; and (b) subject-, object-, and adverbial-clauses which utilize relative pronouns. Included in the word boxes for each item was an additional and unnecessary pronoun (e.g., it, she, him) in order to test whether examinees would be able to construct relative clauses without using resumptive pronouns (a widely attested interlanguage phenomenon). Item 73 below provides an example.

Item 73:

Client: “Is this a successful company?”

Words:

owns	a lot	she	earns	who	the business

Stock broker: “*Well, the woman* ___________ ___________ ___________ .”

***Section** 7*. Finally, a ten-item cloze procedure was included in the GP1 pilot test. This cloze test had been developed (Brown, 1993) in order to provide a quick estimation of general English language ability differences among students at Japanese universities. It was included in the current investigation in order to provide an additional criterion, besides the participant groupings based on language program placements, for distinguishing among participants in terms of global language ability differences. It was hoped that results on the cloze test would provide an additional basis for comparison with outcomes on the pilot test items. Examinees were instructed to first read through the cloze passage and then to provide a single word making the most sense for each of ten blanks in the passage. An exact-response approach, based on native-speaker baseline performances, was used to score examinee responses.

Initial versions of all sections of the GP1 were pilot tested with three individuals who provided feedback on the clarity of instructions and the appropriateness of individual items. Their suggestions for revision were incorporated and the operational pilot test instrument was compiled, printed, and copied for pilot administration.

Test Administration and Scoring Procedures

Administration. The pilot test GP1 was administered to 57 participants in several sessions over the course of one week. Participants were divided during recruitment into the following three testing groups, such that examinees of relatively similar language proficiency levels were administered the test in the same testing session: (a) participant level-groupings 1 and 2; (b) participant level-grouping 3; and (c) participant level-groupings 4 and 5. These efforts were made in order to ensure that instructions could be adequately explained to and understood by all pilot examinees. Because examinees were from different first language backgrounds and different English proficiency levels, several additional steps were taken in order to make sure that examinees understood what was expected of them on all items.

The test proctor carefully delivered all instructions and controlled participants' progression from the beginning to the end of each testing session (full test administration instructions are provided in Appendix B). After all examinees had been admitted to a given session, the proctor handed out test forms and pencils, instructing examinees not to begin until he told them to do so. He then read the consent form statement aloud and clarified any questions about the form; examinees were then asked to sign the form if they agreed to give their consent. The proctor proceeded to read the general test instructions aloud, while examinees read along on their test forms, and he again clarified any questions. When all examinees had completed the background information form following these general instructions, the proctor instructed them to turn to page 4 of the test and to read the instructions for Section 1 while he read them aloud. After checking for any clarification requests, he proceeded to explain the item format with the help of the example item. He also demonstrated the importance of supplying only one word per blank (for these fill-in items) by modeling the process on a chalkboard (and these careful explanations were apparently successful, as none of the 57 examinees supplied more than one word for any one blank). Once he had ascertained that all examinees understood the item format, he instructed them to work through items 1-8 and to make sure that they did not proceed beyond item 8 until he told them to do so. While examinees worked, the proctor monitored visually in order to make sure that they stopped on page 6 of the test.

Once all examinees had completed Section 1 of the test, the proctor instructed them to proceed to page 7. Once again, he carefully read instructions for Section 2 aloud, worked through the example item, and answered any item format questions. He also carefully explained that words presented in the item prompts within a single box were to be kept together when filling in the item response blanks, and he demonstrated this process on the chalkboard (again, this careful process resulted in all of the examinees keeping words from a single box together when they re-wrote them in the response blanks for all items using this format). Examinees were again monitored until they had all finished the section, and the proctor made sure that no examinees returned to the previous section or proceeded to a subsequent section. Similar careful administration procedures were followed for Sections 3-7, with the proctor highlighting item format peculiarities for each section as necessary. After examinees had finished the exam, the proctor collected all test forms and proceeded to distribute participant compensation (in the form of one ten-dollar bill per participant). He also requested that participants sign and date forms acknowledging receipt of payment.

Scoring. Although only constrained constructed-response formats were used on the test, thereby minimizing the range of responses for any one item, several possibilities nevertheless presented themselves for scoring individual items on Sections 2 through 6 of the test (i.e., the word order items of interest for operational test use). Each of these scoring possibilities depended in turn on the eventual interpretations to be based on a given item and the final full-test collection of items. One possibility was to pursue an “interlanguage-sensitive” approach to scoring. This approach would enable test performances to serve as evidence for interpretations about the *emergence* of a particular developmental stage within a learner’s developing knowledge of word order rules in English (as discussed above). Thus, for a single item that tested a given word order rule, all possible word order combinations available for the item would be determined a priori as providing evidence for a particular stage within the posited developmental hierarchy, and item responses would be scored accordingly. For example, a single item testing question formation in English might provide evidence that a learner was at any one of three possible stages, depending on the combination of words produced by the learner. A second possibility for scoring items in Sections 2-6 on the GP1 was to adopt a “target-like accuracy" approach. This approach would enable performances on individual items to be interpreted as either target-like or not (i.e., right or wrong) according to word order norms for standard English. Thus, for an item testing a given phenomenon in English word order, only the target-like word order combination would be identified a priori, and any deviation from this target would be scored as incorrect (i.e., regardless of the fact that the range of possible deviations might provide evidence for interpretations about developmental stages preceding target-like levels of L2 processing).

Several factors argued against the use of interlanguage-sensitive scoring for the purposes of the current project. First, the primary purpose for GP1 was to identify a small set of English grammar items which could be used to separate potential examinees into three broad ability groups, each of which would be predicted to process language *on the whole* in increasingly target-like ways. A few items representing several rules of English word order would therefore be used to inform interpretations about very broad global differences among examinees. Scoring and interpreting these few items in interlanguage-sensitive ways would require of a given item that it tap multiple constructs (i.e., multiple developmental stages), and that it do so for each construct with high degrees of consistency for all examinees. Unfortunately, while research has shown that the particular word order rules in question *emerge* in implicational ways within a learner's developing mental grammar (i.e., the order in which learners begin to process each rule is highly predictable), research has also shown that individual language learners are extremely variable in terms of the linguistic and communicative contexts in which they initially apply a given rule. Thus, for current testing purposes, scoring only 12 items in interlanguage-sensitive ways would likely result in highly variable patterns of performance, since only a very finite set of linguistic and communicative contexts could be provided. Intra-individual variation across these items would most likely result in non-interpretable results. For warranted interpretations to be drawn about the emergence of a given word order developmental stage, many more items (covering many more linguistic and communicative contexts) would be required.

A second problem with the use of interlanguage-sensitive scoring for current testing purposes is that, while developing target-like levels of accuracy with a given word order rule may require extensive opportunities for language learning and use, research findings suggest that initial *emergence of the full range* of predicted stages in word order development may occur very early among instructed language learners (Ellis, 1989; Hudson, 1993; Norris, 1996). Therefore, for the purposes of separating examinees into three groups whose language processing abilities differ broadly, a focus on *emerging* levels of language processing (in this case, the processing of English word order rules) would not prove particularly useful.

A final, mechanical problem is that interlanguage-sensitive scoring would be a much more tedious process than target-like accuracy scoring, requiring separate scoring criteria for each item and multiple scorers for the current investigation (or a much more complex scoring algorithm for computer-based versions of the test).

For these reasons, and in order to provide for a more stable estimation of broad language processing ability differences among examinees, the "target-like accuracy" approach was adopted for scoring individual items on the GP1. Each item was therefore developed with a single combination of words (based on norms from standard English) serving as the target, and all items were reviewed for alternative feasible word order combinations (and revised accordingly). Item responses were scored by hand, with any deviation from the target scored as incorrect.

Data Analyses

To summarize the current investigation, 57 volunteer participants completed the eight morphology items, the 67 pilot word order items, and the ten-item cloze test. Item responses were scored for target-like accuracy, and all data were entered into a Microsoft Excel spreadsheet for further analysis. The following analyses were conducted on pilot test data: (a) a multi-faceted Rasch model analysis (Linacre, 1998) was conducted for examinee performances on items in Sections 2-6 of the test; (b) descriptive statistics were calculated for all measures of examinee ability and all item types; (c) correlational analyses were conducted comparing all measures of examinee ability; (d) one univariate analysis of variance (ANOVA) procedure was conducted to compare measures of examinee ability among the five participant groupings; (e) classical test theory item analyses, including item facility and item discrimination calculations, were conducted;

and (f) descriptive statistics, correlations, reliability and decision dependability indices were estimated for several new test forms compiled from pilot test items.

RESULTS AND DISCUSSION

This section presents and discusses key results, in the following order: (a) full-test and sub-section outcomes; (b) item attributes; and (c) item revisions and sampling strategies for operational test forms.

Full-Test and Sub-Section Outcomes

A multi-faceted Rasch model analysis using FACETS (Linacre, 1998) was conducted for the subset of English word order items in Sections 2-6 (i.e., the items being piloted for eventual operational test use). The dichotomous model explored two facets, examinees and items, with convergence criteria set as a maximum difference between expected and actual score points of 0.10 raw score points for all elements and a maximum change of 0.01 logits in the previous iteration. Convergence was achieved after 23 iterations and the connection of the subsets was verified. Outcomes showed that the model was moderately able to separate examinees into distinct ability levels (person separation index = 2.08, corresponding to a reliability of 0.81) and to distinguish among item difficulties (item separation index = 2.78, corresponding to a reliability of 0.89). However, as shown in Figure 1, while the model indicated broad ability differences among examinees (from –2.80 to 3.24 logits) and broad difficulty differences among items (from –6.18 to 2.26 logits), it also showed that the difficulty of many test items fell well below the lowest examinee ability levels. Note that, in Figure 1, examinee ability estimates are indicated in the second column, with each star symbol representing one examinee; examinee ability increases moving from the bottom to the top of the figure. Likewise, item difficulty estimates are represented in the third column of Figure 1, with item difficulty increasing from the bottom to the top of the figure (for much more on the interpretation of FACETS analyses, see McNamara, 1996). It is apparent that, while examinees fell into a relatively normal distribution, item difficulties were skewed, with most bunched towards the bottom of the table, and a handful of items stretching up the scale to just over 0.50 logits (and two items extending to just over 2.00 logits). These initial results reflect the fact that most of the examinees performed well on most of the word order items and that only a subset of the items would seem useful for distinguishing among examinee abilities. This issue will be addressed again below in the discussion of item attributes.

```
------------------------------
|Measr|+Examinee|-Items      |
------------------------------
+   4 +         +           +
|     |         |           |
|     |         |           |
|     |         |           |
|     | *       |           |
+   3 +         +           +
|     |         |           |
|     |         |           |
|     | ***     |           |
|     |         | *         |
+   2 +         +           +
|     | *****   |           |
|     |         |           |
|     | ****    |           |
|     |         |           |
+   1 + ****    +           +
|     |         |           |
|     | ******  | *.        |
|     | **      |           |
|     | *****   |           |
*   0 * **      * *         *
|     |         | .         |
|     | *****   | .         |
|     | **      | .         |
|     | *       | .         |
+  -1 + *       + .         +
|     | ****    |           |
|     | ***     | *         |
|     | *       | .         |
|     |         |           |
+  -2 + ***     + *.        +
|     |         | *         |
|     | **      | .         |
|     | *       | .         |
|     | **      | ***.      |
+  -3 +         + .         +
|     |         |           |
|     |         | *         |
|     |         |           |
|     |         | ***.      |
+  -4 +         +           +
|     |         | ****      |
|     |         |           |
|     |         |           |
|     |         |           |
+  -5 +         + ********** +
------------------------------
|Measr| * = 1   | * = 2,. = 1|
------------------------------
```

Figure 1. FACETS measure estimates of examinee ability and item difficulty on a common scale

Descriptive statistics were calculated for examinee scores on the full test, as well as for the "Mike" cloze passage, the test sub-sections, the word order sub-sections 2-6 together, and the FACETS measure outcomes. Means and standard deviations in Table 4 show that examinees scored consistently well on most of the test sub-sections as well as on the full test and on the word order sections combined. Skew statistics also demonstrate that most of the sub-sections produced negatively skewed distributions of scores for this population sample. These results add support to the interpretation above that the majority of the test items were correctly answered by most of the pilot test examinees. Among the operational test sections, the only exception to this pattern was observed in Section 6 (the relative clause items), where mean examinee scores were well centered and where virtually no skew was observed in the score distribution. These items were apparently more difficult than items in other sections for many of the examinees; the Section 6 items may prove particularly useful in distinguishing among broad examinee ability groups. Scores on the cloze passage were well centered and showed virtually no skew, and examinees scored at the entire range of possible scores. This finding suggests that examinee language abilities did differ within the participant sample, at least as measured by this brief test.

Table 4

Descriptive Statistics for Full-Test and Sub-Section Outcomes

Statistic	Cloze	S1	S2	S3	S4	S5	S6	Total	S2-S6	FACETS Examinee
K	*10*	*8*	*15*	*9*	*30*	*7*	*6*	*75*	*67*	*67*
Mean	5.60	5.16	14.21	8.25	26.28	5.61	2.51	62.02	56.86	0.00
SD	*2.46*	*1.81*	*1.11*	*1.42*	*2.35*	*1.96*	*2.16*	*8.18*	*6.89*	*1.46*
Min	1.00	1.00	11.00	3.00	18.00	0.00	0.00	41.00	40.00	-2.80
Max	10.00	8.00	15.00	9.00	30.00	7.00	6.00	74.00	66.00	3.24
Skew	-0.08	-0.64	-1.32	-2.13	-1.37	-1.33	0.09	-0.84	-0.93	-0.08

In order to further investigate the language ability differences among participants, mean scores were calculated for each participant level-grouping (based on ESL program levels), and three univariate analysis of variance (ANOVA) procedures were conducted in order to compare average group performances on three dependent variable measures. An overall statistical decision level was set at $p < 0.05$ for ANOVA comparisons; because three comparisons were conducted on the same population sample, the alpha level was adjusted to $p < .01$, to decrease the chance of spurious findings of group differences. The ANOVA procedures compared differences between the five participant level-groups (the independent variable) for each of three dependent variables: the Mike test, the total test score, and the total score for word order test Sections 2-6. Main differences for each dependent variable (F = 10.03, 27.62, 22.81, respectively) were statistically significant ($p = .001$ for each). Scheffe post-hoc comparisons between the five participant level-groups revealed the following patterns of statistically significant group difference for each of the three dependent variables: (1) < (2) < (3) = (4) = (5). As is clearly shown in Table 5, large mean differences on the three variables were found between groups (1) and (2), as well as between each of these groups and groups (3), (4), and (5). Furthermore, mean performance outcomes on each dependent variable for each of the top three participant level-groups were found to be very similar, not only for the three summary variables, but also for performances on each test sub-section. Such similar performances by examinees at the three top program level-groups suggest that average language ability differences may not have differed very much among these participant groups.

Table 5 also shows that the most substantial differences in performance on word order items between the lower two and the upper three participant level-groups were found for Sections 5 and 6, while Sections 2, 3, and 4 revealed only minimal differences between group (1) and the other groups. These findings suggest that word order items in sections 5 and 6 might prove the most useful for distinguishing among very broad examinee ability differences.

Table 5

Mean Test Scores by Participant Level-Group

										FACETS
Group	Cloze	S1	S2	S3	S4	S5	S6	Total	S2-S6	Examinee
(k)	*10*	*8*	*15*	*9*	*30*	*7*	*6*	*75*	*67*	*67*
(1)	3.10	2.90	13.10	6.90	23.60	3.40	0.00	49.90	47.00	-1.87
(2)	4.63	4.74	14.00	8.11	25.84	5.37	1.63	59.68	54.95	-0.58
(3)	7.10	5.90	15.00	8.90	27.30	6.90	3.90	67.90	62.00	1.22
(4)	7.22	6.33	14.67	9.00	27.44	6.22	4.22	67.89	61.56	0.98
(5)	7.11	6.56	14.56	8.56	27.89	6.56	3.89	68.00	61.44	0.96

In order to further investigate apparent relationships between test scores and participant abilities, Pearson product-moment correlation coefficients were calculated among all pairs of test score variables. Table 6 shows moderately strong correlations between participant ESL program level-group (LEVEL) and total GP1 test scores as well as total scores on the word order items in Sections 2-6. Scores on Section 6 word order items also showed a similar relationship with participant level. Pilot test performance outcomes thus seem to be somewhat related with program level placement. Note that the "Mike" cloze passage also correlated only moderately with both program level and with scores on the pilot test. These findings suggest that, while average language ability differences exist among participants at the different ESL program levels, these differences are certainly not categorical for individuals at each level. Correlation coefficients calculated for the subset of examinees who reported TOEFL scores also revealed only moderate relationships between this variable and program level-group (0.68), and between TOEFL scores and total scores on word order items in Sections 2-6 (0.46). Apparently, while examinee language abilities at the different program levels differ somewhat on average, there is also extensive variability within and across program levels, and especially at level-groups (3), (4), and (5).

Table 6

Pearson Correlation Coefficients for Pilot Test Scores

	LEVEL	Cloze	S1	S2	S3	S4	S5	S6	TOTAL	S2-S6	FACETS
											Exam.
LEVEL	1.000	.595	.651	.435	.392	.572	.483	.659	.725	.690	.679
MIKE		1.000	.533	.489	.250	.243	.393	.707	.578	.546	.527
S1			1.000	.461	.444	.543	.450	.542	.768	.649	.661
S2				1.000	.521	.420	.455	.483	.685	.693	.636
S3					1.000	.327	.529	.338	.652	.658	.611
S4						1.000	.620	.479	.796	.803	.769
S5							1.000	.508	.805	.837	.774
S6								1.000	.767	.768	.815
TOTAL									1.000	.986	.960
S2-S6										1.000	.966
FACETS											1.000

Note. All correlations statistically significant, $p < .05$, $N = 57$.

Item Attributes

In order to investigate the extent to which word order items were functioning as intended, as well as to motivate item revision and sampling for operational testing purposes, several classical test theory indices of item quality were calculated for each item. Results of these item analyses are displayed along with FACETS item measure statistics for each item in Table 7. The attributes of GP1 test items will first be discussed on the basis of item analysis results across the focal word order sub-sections of the test (2-6) and subsequently for each test section.

Item Facility (IF) indices were calculated on each item for all examinees, for each participant level-group, and for the lower/middle/upper thirds of examinees (based on total scores from section 2-6 combined). IF indices show the proportion of examinees who answered a given item correctly. Overall, IF values ranged from a low of 0.16 (Item 71) to a high of 1.00 (numerous items), and generally high IF values indicate that many of the word order items were answered correctly by most or all of the examinees. Where differences existed in examinee performances on particular items, these differences tended to be reflected in the lowest IF indices for participant group (1) and the lower third of scorers, and in somewhat lower scores by participant group (2) and the middle third of scorers. Consistent differences among the upper three participant level-groups (3, 4, and 5) were not apparent for any of the items. However, a handful of items did reveal substantial IF differences between the upper third of examinees and the bottom two-thirds of examinees.

In order to further explore the extent to which individual items distinguished among examinees, *Item Discrimination* (ID) indices were also calculated. ID simply reflects the difference between IF values for two groups of examinees, and it provides one basis for sampling items according to the extent to which they discriminate among apparent examinee ability differences. For the current project, ID was calculated between the top third and the middle and lowest thirds of examinees (thirds based on total score on Sections 2-6 of the test), as well as between the middle third and the lowest third of examinees. With several exceptions, ID indices revealed very little discriminating power for items on Sections 2, 3, and 4 of the test, while Sections 5 and 6 contained a number of items which discriminated much better between low-, middle-, and high-scoring examinees. These findings are further examined on a section-by-section basis below.

The last three columns in Table 7 show item measure statistics from the FACETS analysis reported above. The item measure itself is an index of item difficulty, which can be directly compared with examinee ability estimates on the same scale (see Figure 1 above). Negative item measure values indicate items that fell below the mean ability level of examinees, while positive values indicate items that fell above the mean examinee ability. Standard error (S.E.) values show the amount of error involved in interpretations based on a given item (i.e., the lower the better), and point-biserial correlation coefficients (r_{pbi}) show the strength of relationship between a single item and the total test score (i.e., the higher the better). Generally, FACETS item analyses supported the patterns observed in the classical test theory item analyses above, showing particularly easy items with low discriminating power and higher error on Sections 2-4 of the test, and increasing numbers of more difficult, higher discriminating, and lower error items on the final two word order sections. These patterns are explored in more detail for each test section below.

Before proceeding, it should be noted that one additional set of analyses was undertaken for all examinee performances on all items. In order to investigate whether or not there might be an *implicational* basis for interpreting examinee abilities, based on performances observed in the current project, individual examinee performances were examined on all items representing each of the word order rule stages. No categorical relationships were found among word order rules at the different stages in the current test data. That is, every examinee who took the pilot test answered correctly at least one item on every section and for every word order rule in the test, with the single exception of the relative clause items, where a number of examinees missed all of the items.

Table 7

Item Attributes

Item #	Item type	IF total	IF partip group 1	IF partip group 2	IF partip group 3	IF partip group 4	IF partip group 5	IF lower 1/3	IF mid 1/3	IF upper 1/3	ID mid-low	ID hi-low	ID hi-mid	Facets item measure	*S.E.*	r_{pbi}
9	CAN	0.96	0.90	1.00	1.00	1.00	0.89	0.94	0.95	1.00	0.01	0.06	0.05	-4.18	0.75	0.21
10	CAN	0.98	1.00	0.95	1.00	1.00	1.00	0.94	1.00	1.00	0.06	0.06	0.00	-4.93	1.04	0.23
11	CAN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18	1.83	0.00
12	CAN+N	0.95	1.00	0.89	1.00	0.89	1.00	0.94	0.95	0.94	0.01	0.00	-0.01	-3.72	0.69	-0.03
13	CAN+N	0.96	0.80	1.00	1.00	1.00	1.00	0.89	1.00	1.00	0.11	0.11	0.00	-4.18	0.78	0.33
14	CAN+N	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18	1.83	0.00
15	CAN+N	0.88	0.50	0.89	1.00	1.00	1.00	0.67	0.95	1.00	0.29	0.33	0.05	-2.64	0.45	0.42
16	CAN+N	0.98	0.90	1.00	1.00	1.00	1.00	0.94	1.00	1.00	0.06	0.06	0.00	-4.93	1.04	0.23
17	CAN+N	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18	1.83	0.00
18	CAN+N	0.89	0.80	0.79	1.00	1.00	1.00	0.67	1.00	1.00	0.33	0.33	0.00	-2.85	0.48	0.37
19	CAN+N	0.89	0.70	0.89	1.00	1.00	0.89	0.72	0.95	1.00	0.23	0.28	0.05	-2.85	0.52	0.49
20	CAN+N	0.96	0.80	1.00	1.00	1.00	1.00	0.89	1.00	1.00	0.11	0.11	0.00	-4.18	0.78	0.11
21	CAN+N	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18	1.83	0.00
22	CAN+N	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18	1.83	0.00
23	CAN+N	0.74	0.70	0.58	1.00	0.78	0.78	0.56	0.77	0.88	0.22	0.33	0.11	-1.43	0.41	0.23
24	X INV	0.81	0.50	0.79	1.00	1.00	0.78	0.50	0.91	1.00	0.41	0.50	0.09	-1.96	0.40	0.49
25	X INV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18	1.83	0.00
26	X INV	0.81	0.50	0.79	0.90	1.00	0.89	0.61	0.82	1.00	0.21	0.39	0.18	-1.96	0.39	0.41
27	X INV	0.96	1.00	0.89	1.00	1.00	1.00	0.89	1.00	1.00	0.11	0.11	0.00	-4.18	0.74	0.25
28	X INV	0.95	0.80	0.95	1.00	1.00	1.00	0.83	1.00	1.00	0.17	0.17	0.00	-3.72	0.67	0.43
29	X INV	0.95	0.90	0.89	1.00	1.00	1.00	0.83	1.00	1.00	0.17	0.17	0.00	-3.72	0.62	0.31
30	X INV	0.89	0.60	0.89	1.00	1.00	1.00	0.67	1.00	1.00	0.33	0.33	0.00	-2.85	0.51	0.51
31	X INV	0.93	0.70	0.95	1.00	1.00	1.00	0.78	1.00	1.00	0.22	0.22	0.00	-3.37	0.56	0.35
32	X INV	0.95	0.90	0.95	1.00	1.00	0.89	0.89	0.95	1.00	0.07	0.11	0.05	-3.72	0.66	0.15

Table 7

Item Attributes (continued)

Item #	Item type	IF total	IF partip group 1	IF partip group 2	IF partip group 3	IF partip group 4	IF partip group 5	IF lower 1/3	IF mid 1/3	IF upper 1/3	ID mid-low	ID hi-low	ID hi-mid	Facets item measure	*S.E.*	r_{pbi}
33	Y-N INV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18	1.83	0.00
34	Y-N INV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18	1.83	0.00
35	Y-N INV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18	1.83	0.00
36	Y-N INV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18	1.83	0.00
37	Y-N INV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18	1.83	0.00
38	Y-N INV	0.96	1.00	1.00	1.00	0.78	1.00	1.00	0.95	0.94	-0.05	-0.06	-0.01	-4.18	0.81	-0.13
39	Y-N INV	0.98	0.90	1.00	1.00	1.00	1.00	0.94	1.00	1.00	0.06	0.06	0.00	-4.93	1.08	0.29
40	Y-N INV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18	1.83	0.00
41	Y-N INV	0.96	1.00	0.95	1.00	0.89	1.00	1.00	0.91	1.00	-0.09	0.00	0.09	-4.18	0.81	-0.06
42	Y-N INV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18	1.83	0.00
43	Y-N INV	0.96	0.90	0.95	1.00	1.00	1.00	0.94	0.95	1.00	0.01	0.06	0.05	-4.18	0.77	0.10
44	Y-N INV	0.84	0.70	0.74	0.90	1.00	1.00	0.61	0.95	0.94	0.34	0.33	-0.01	-2.28	0.42	0.39
45	Do-aux 2nd	0.93	0.80	0.89	1.00	1.00	1.00	0.83	0.95	1.00	0.12	0.17	0.05	-3.37	0.61	0.12
46	Do-aux 2nd	0.98	0.90	1.00	1.00	1.00	1.00	0.94	1.00	1.00	0.06	0.06	0.00	-4.93	1.02	0.17
47	Do-aux 2nd	0.89	0.80	0.95	0.90	0.89	0.89	0.89	0.82	1.00	-0.07	0.11	0.18	-2.85	0.50	0.23
48	Do-aux 2nd	0.96	0.90	0.95	1.00	1.00	1.00	0.89	1.00	1.00	0.11	0.11	0.00	-4.18	0.76	0.28
49	Do-aux 2nd	0.95	0.90	0.89	1.00	1.00	1.00	0.83	1.00	1.00	0.17	0.17	0.00	-3.72	0.64	0.35
50	Do-aux 2nd	0.95	0.80	0.95	1.00	1.00	1.00	0.89	0.95	1.00	0.07	0.11	0.05	-3.72	0.67	0.36
51	Do-aux 2nd	0.86	0.70	0.84	0.90	0.89	1.00	0.72	0.91	0.94	0.19	0.22	0.03	-2.45	0.46	0.23
52	Do-aux 2nd	0.89	0.70	0.89	1.00	1.00	0.89	0.78	0.91	1.00	0.13	0.22	0.09	-2.85	0.49	0.41
53	Do-aux 2nd	0.98	0.90	1.00	1.00	1.00	1.00	0.94	1.00	1.00	0.06	0.06	0.00	-4.93	1.09	0.31
54	Do-aux 2nd	0.89	0.80	0.84	0.90	1.00	1.00	0.78	0.91	1.00	0.13	0.22	0.09	-2.85	0.49	0.26
55	Do-aux 2nd	0.98	0.90	1.00	1.00	1.00	1.00	0.94	1.00	1.00	0.06	0.06	0.00	-4.93	1.05	0.08
56	Do-aux 2nd	0.68	0.40	0.53	0.80	0.89	1.00	0.33	0.77	0.94	0.44	0.61	0.17	-1.07	0.35	0.53
57	Do-aux 2nd	0.91	0.80	0.95	0.90	1.00	0.89	0.83	0.95	0.94	0.12	0.11	-0.01	-3.09	0.51	0.25

Table 7

Item Attributes (continued)

Item #	Item type	IF total	IF partip group 1	IF partip group 2	IF partip group 3	IF partip group 4	IF partip group 5	IF lower 1/3	IF mid 1/3	IF upper 1/3	ID mid-low	ID hi-low	ID hi-mid	Facets item measure	*S.E.*	r_{pbi}
58	Tag Q	0.54	0.20	0.58	0.50	0.78	0.67	0.28	0.59	0.76	0.31	0.49	0.17	-0.23	0.35	0.38
59	affirm-neg Q	0.16	0.00	0.11	0.40	0.11	0.22	0.06	0.18	0.24	0.13	0.18	0.05	2.26	0.44	0.18
60	Y-N INV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18	1.83	0.00
61	Tag Q	0.60	0.40	0.58	0.50	0.89	0.67	0.33	0.68	0.76	0.35	0.43	0.08	-0.54	0.36	0.35
62	affirm-neg Q	0.39	0.20	0.26	0.60	0.33	0.67	0.28	0.41	0.47	0.13	0.19	0.06	0.67	0.39	0.06
63	X INV+ Q	0.89	0.70	0.84	1.00	1.00	1.00	0.67	1.00	1.00	0.33	0.33	0.00	-2.85	0.51	0.29
64	X INV+ Q	0.95	0.90	0.89	1.00	1.00	1.00	0.83	1.00	1.00	0.17	0.17	0.00	-3.72	0.63	0.35
65	X INV+ Q	0.65	0.20	0.58	0.90	0.78	0.89	0.28	0.73	0.94	0.45	0.66	0.21	-0.85	0.35	0.54
66	X INV+ Q	0.75	0.40	0.74	1.00	0.78	0.89	0.39	0.86	1.00	0.47	0.61	0.14	-1.55	0.41	0.65
67	X INV+ Q	0.72	0.30	0.68	1.00	0.78	0.89	0.33	0.82	1.00	0.48	0.67	0.18	-1.31	0.40	0.66
68	X INV+ Q	0.84	0.50	0.84	1.00	1.00	0.89	0.56	0.95	1.00	0.40	0.44	0.05	-2.28	0.47	0.59
69	X INV+ Q	0.81	0.40	0.79	1.00	0.89	1.00	0.44	0.95	1.00	0.51	0.56	0.05	-1.96	0.45	0.65
70	Relative clause	0.40	0.00	0.26	0.60	0.67	0.67	0.06	0.36	0.82	0.31	0.77	0.46	0.57	0.35	0.53
71	Relative clause	0.14	0.00	0.21	0.20	0.22	0.00	0.00	0.14	0.29	0.14	0.29	0.16	2.26	0.42	0.28
72	Relative clause	0.40	0.00	0.21	0.60	0.78	0.67	0.06	0.41	0.76	0.35	0.71	0.36	0.57	0.33	0.48
73	Relative clause	0.51	0.00	0.32	0.80	0.78	0.89	0.11	0.45	1.00	0.34	0.89	0.55	-0.03	0.38	0.63
74	Relative clause	0.56	0.00	0.32	0.90	0.89	1.00	0.11	0.59	1.00	0.48	0.89	0.41	-0.33	0.38	0.65
75	Relative clause	0.49	0.00	0.32	0.80	0.89	0.67	0.11	0.45	0.94	0.34	0.83	0.49	0.07	0.37	0.59

Section 1 items. Several items on Section 1 (morphology items) did seem to discriminate relatively well among examinees, and general patterns suggested that examinees who did not produce evidence of ability to supply the two morphological rules (third person –s and plural –s) also struggled with word order items representing the more advanced developmental stages. However, outcomes on these items are not included for the further purposes of the current project, owing to the inconsistency with which items elicited examinee performances. Thus, while many of the examinees produced the expected lexical items (with or without appropriate morphological marking), a number of examinees produced divergent responses (based on unexpected lexical and morphological forms; e.g., "ate" instead of "eats") which could not be scored relevant to the constructs of interest for this test section. As such, Section 1 outcomes proved too unstable for the purposes of further comparisons with the word order items. Future research should investigate elicitation methods, potentially utilizing different and more controlled item formats, which will enable consistent interpretations to be made about these and related morphological phenomena.

Section 2 items. Section 2 items (CAN, CAN+N) proved, with only a few exceptions, to be uniformly easy and correctly answered by virtually all examinees. This is not surprising, as these items represent very initial stages in word order development, and examinees from instructed settings will certainly have had extensive exposure to related rules. While items from this section generally did not discriminate among examinees at all, there may be a good argument for retaining at least a few for operational testing purposes (as discussed in the item revision and sampling section below). Several CAN+N items did prove to be moderately more difficult, including items 15, 18, 19, and 23. Inspection of items 18, 19, and 23 revealed that these items were answered incorrectly almost always as a result of examinee errors with adverb placement rules in English, but not as a result of the placement of the negator within the sentence. Were these items to be re-scored on the basis of the intended construct (i.e., only the placement of the negator), virtually all examinees would have answered correctly. Given such variability in performance outcomes based on target-like accuracy scoring, these items do not seem appropriate for further consideration. Nevertheless, the performance errors observed with adverb placement rules do suggest a fruitful area for future research. Item 15 discriminated relatively well between a handful of the lowest ability learners and the rest of the pilot test examinees. This item tested post-verbal placement of the negator "no", and where it was answered incorrectly, the answer was always based on preverbal placement of "no" (i.e., the item seemed to be a good test of this particular word order phenomenon). Given this observed consistency, and since this error type reflects an early stage in word order development, this item may prove useful for further testing purposes.

Section 3 items. Items in this section tested the highest posited stage in word order development (although possibly not the highest difficulty item type represented on the current test; see sections 5 and 6 below), cancel inversion (X INV). Several items in Section 3 proved to discriminate relatively well between the overall lowest scoring examinees and all others. In particular, items 24, 26, and 30 showed promise for the purposes of distinguishing the lowest scoring examinees from all others. Closer inspection of these three items revealed that all examinee errors reflected the construct; that is, responses were incorrect because examinees failed to cancel subject-verb inversion in indirect questions. These three items were therefore retained for further consideration.

Section 4 items. Section 4 tested several rules of question formation in English, each of which represented developmental stages prior to those represented by items in Section 3. All Y-N INV items proved very easy for virtually all examinees. However, as noted in the methodology section of the current paper, it is possible that the example item for this test section divulged the correct word order for these items to some examinees. Several observations from the pilot test outcomes speak against this interpretation. First, the inversion rule in question formation reflects a lower developmental stage than other question formation rules tested here, so it would not be unreasonable to expect such high and consistent levels of performance. Second, Do-aux 2nd question formation rules also revealed generally very high IF indices, even though the word order for these items was not modeled in the example item. Third, while most Y-N INV items were answered correctly by all examinees, several items did reveal slightly lower performance levels. As such, it is probably safe to interpret the Y-N INV items as simply having little discriminatory power for the current sample of examinees. Nevertheless, there may be good arguments for retaining several such items for operational testing purposes (as discussed in the item revision and sampling section below).

Several of the Do-aux 2nd items (51, 52, 54) did prove somewhat more discriminatory among the very lowest ability examinees, and all of these items appeared to be testing the intended rule (movement of the auxiliary verb to the position directly following WH question words). All of these items were retained for further consideration as operational test items. Two items in this section showed much higher levels of discrimination (items 44 and 56). However, both of these items involved somewhat infrequent sentence formulations, and as such may not prove appropriate for further consideration. In addition, items from subsequent sections proved to be equally discriminating and more reflective of the average performance levels for a given rule than did these two items.

Finally, the tag-questions and affirmative-negative questions proved to be quite difficult for many examinees. Affirmative-negative questions resulted in very low scores from examinees at all ability levels and in unpredictable variability between participant level-groups, and as such, they were eliminated from further consideration. While the two tag-questions discriminated rather well between the lower third of examinees and all others, they did not discriminate well between the middle and upper scoring groups. What is more, substantial variability was noted in individual performances among examinees in the middle and upper scoring groups, as were generally lower levels of performance among the upper third of examinees. It may have been the case that exposure to or familiarity with tag (as well as affirmative-negative) questions played a primary role in examinee performance outcomes on these items. Given the potential for unpredictability in learner performances under operational circumstances, these item types were precluded from further consideration, although the corresponding word order formulations would be beneficially investigated in future research.

Section 5 items. Items in this section tested the cancel inversion (X INV) rule embedded within questions. On the whole, these items proved to be more difficult than preceding items on the test, and very high ID indices for all but one item suggested that they were quite effective at distinguishing between lower level examinees and middle and upper level examinees, with the upper two-thirds performing these items with generally high degrees of success. Because items also appeared to be testing the construct (i.e., examinees missed these items because they did not cancel subject-verb inversion in indirect questions), items 63, 65, 66, 67, 68, and 69 were therefore retained for further consideration. Item 64 proved to have very low discriminatory power and was eliminated.

Section 6 items. Section 6 items tested relative clause formation and had been posited to be the most difficult items on the pilot test GP1. All six of these items showed much lower IF values than did items on preceding sections of the test (with the exception of the experimental tag and affirmative-negative question formation items), and five of the six items revealed very high item discrimination indices between all three examinee levels (low, middle, and high). These relative clause items were the only items on the test to show substantial ID values between the highest scorers on the test and the middle scorers on the test, a good indication that these items would serve well for distinguishing advanced from other examinees. What is more, examinees from participant level groupings 1 (IF = 0.00 for all items) and 2 (IF ranged from 0.21 to 0.33 for all items) scored very poorly on these items. In light of these findings, all items from Section 6 were retained for further consideration, with the exception of item 71, which proved too difficult for even the top scoring examinees.

Item Revision and Sampling for Operational Test Forms

Based on results of the test and item analyses, there are several feasible approaches to revising items and sampling the three 12-item test forms required in order to fulfill operational testing purposes. Which of these approaches is pursued will depend largely on the extent to which the population of participants sampled in the current investigation, and their patterns of performance on the pilot items, may be assumed to adequately represent the target population to be served in the operational web-based testing context. Unfortunately, until sufficient demographic information is available from the test-use context, decisions regarding the appropriate constellation of test items will remain speculative. Nevertheless, demographic evidence from the pilot test investigation may at least offer a useful point of departure in this endeavor.

It should be recalled that all participants in this pilot test investigation had experienced some English language instruction, and all were enrolled in English classes when they took the test. In addition, many of the examinees (i.e., those in level-groups 3, 4, and 5) had already exhibited English language proficiencies sufficient to be admitted to courses of study at a U.S. university (albeit with the caveat that they engage in further English language training), with average TOEFL scores ranging between 500 and 600. In light of these observations, it may be the case that the current participant sample did not adequately represent a low enough ability range to be reflective of the likely target population of learners (although participants were recruited from the lowest available L2 English instruction contexts), whereas it probably is the case that participants adequately represented the highest likely ability ranges of target learners (i.e., L2 users of English matriculated in US university programs of study). Test users/developers will need to give careful consideration to this issue when deciding on the final test items and instruments. In order to provide further guidance for this decision, two alternatives for sampling three 12-item test forms are described below. Prior to discussing specific item recommendations for constituting operational test forms, a few overarching considerations are addressed.

Selecting appropriate test items. Several judgments were made in order to sample items for the creation of three new and parallel test forms consisting of 12 items each. It was reasoned that only two decision cut-scores would be necessary on the test; that is, the 12 items selected would only need to distinguish between Beginner and Intermediate examinees and between Intermediate and Advanced Examinees. Relevant cut-scores could therefore feasibly fall at scores ranging from 0 to 12 points (assuming dichotomous scoring). For example, it would be possible to make the Beginner/Intermediate distinction by creating a test with only those items that distinguish between Intermediate and Advanced learners, if the cut-score for Beginners were set at 0 points. Likewise, the Advanced distinction could simply be made at a cut-score of 12 points. However, given uncertainties about the eventual population of examinees for the operational test, it was decided that cut-scores for both decisions should probably fall *within* the range of possible scores, such that items which distinguish well between both Beginner/Intermediate learners and Intermediate/Advanced learners could be included.

In order to identify these operational test items, item constructs and Item Facility, Item Discrimination, and Facets item measure indices were considered in the following manner for all items that had been retained from pilot test investigations. First, it was decided that the full range of word order constructs should be represented; this would enable further investigation of the theoretical foundations for the exam, and it would allow for the possibility that a generally lower ability range of examinees might be encountered under operational test use. Items therefore needed to be sampled from all of the word order rule test sections (S2-S6).

Next, items were needed which would distinguish among the range of examinee abilities actually estimated in the current project. Facets item measure values were therefore utilized to identify items which ranged from relatively high difficulty (0.57 logits) to relatively low difficulty (-2.85 logits) and at various points in between, based on the observed abilities of examinees in the pilot test sample. Recall that examinee abilities, as estimated according to FACETS measures, ranged from –2.80 to 3.24 logits; thus, a number of examinees (approximately one-third) were estimated to have abilities substantially higher than the highest difficulty items available from the pilot test investigations. The lack of more difficult items would prove problematic, were finer-grained distinctions among more advanced English learners to be made on the basis of the operational test. However, because the test only needed to distinguish between an advanced group of learners and an intermediate group of learners (at the highest of the two decision cut-scores), higher difficulty items were unnecessary for current testing purposes (i.e., items appropriate for this decision needed to be of a difficulty roughly equivalent to examinee ability estimates at the transition between intermediate and advanced learners, but no higher).

Next, items were needed which would discriminate well and consistently among three broadly differing groups of examinees. Sampling needed to include items which had high ID values (0.40 and above) for differences between both upper and middle scorers and between middle and lower scorers on the test, such that consistent distinctions could be made between three broad groups. In addition, items were sought which showed relatively low standard error values (0.50 and below, from FACETS analyses) and relatively high item-total (point-biserial) correlations (0.50 and above). Items were also sought which showed consistent increases in IF values from participant level-group 1 to 2, and from level group 2 to 3/4/5, because these distinctions most closely approximated three broadly differing ability groups in the current pilot-test population sample (according to all measures employed in this investigation). It was also reasoned that several further items with very high IF values should be included, in order to introduce examinees to the unique item response format on the test and because they would prove appropriate for Beginner examinees by providing some opportunity for success.

A final consideration for sampling items into three new test forms was that the three forms needed to exhibit parallel difficulties and discriminatory power, such that consistently the same placement decision would be made on the basis of each 12-item test. Where possible, for each word order rule construct, items were therefore sampled such that exactly the same or very similar IF, ID, and Facets measure values would be reflected. Where three such equivalent items did not exist, new items were created or similar existing items were revised in order to replicate as closely as possible the linguistic context and word order rule construct for a selected item. In addition, one of every three items selected to represent a particular construct needed to also reflect business content insofar as the communicative context was concerned (i.e., in order to compile one 12-item test form for the Business English section of the assessment battery).

Sampling approach #1. The first approach to compiling three 12-item test forms was based on the assumption that the pilot participant population sample probably did not adequately represent a lower ability range of examinees who will be tested in operational contexts. As such, a number of very high facility items were included. Tables 8, 9, and 10 show the 12 sampled and revised/new items for each of three new test forms (the three test forms are reproduced in their entirety in Appendix C). Two items on each form represent each of the six word order rule categories investigated in the current project. Note also that instructions for the "ask a question" section were revised such that neither of the word order patterns corresponding to the two question formation rules is modeled in the example. Items range in difficulty from the very easy to the relatively difficult, as reflected by overall IF and Facets item measure values. Items representing each rule also reflect the same or very similar difficulties across the three forms. New or revised items are shown in italicized font to indicate that corresponding item attributes are hypothesized on the basis of the item that served as a model, and the corresponding model item number is shown in parentheses in the second column of each table.

Table 8

Revised Grammar Ability Finder GP2: Form 1

Item #	Item #	Item type	IF	IF	IF	IF	IF	IF	IF	IF	IF	ID	ID	ID	Facets
(new)	(old)		total	partip	partip	partip	partip	partip	lower	mid	upper	mid-low	hi-low	hi-mid	item
				group 1	group 2	group 3	group 4	group 5	1/3	1/3	1/3				measure
1	35	Y-N INV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18
2	33	Y-N INV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18
3	52	Do-aux 2nd	0.89	0.70	0.89	1.00	1.00	0.89	0.78	0.91	1.00	0.13	0.22	0.09	-2.85
4	51	Do-aux 2nd	0.86	0.70	0.84	0.90	0.89	1.00	0.72	0.91	0.94	0.19	0.22	0.03	-2.45
5	21	CAN+N	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18
6	15	CAN+N	0.88	0.50	0.89	1.00	1.00	1.00	0.67	0.95	1.00	0.29	0.33	0.05	-2.64
7	*New(24)*	*X INV*	*0.81*	*0.50*	*0.79*	*1.00*	*1.00*	*0.78*	*0.50*	*0.91*	*1.00*	*0.41*	*0.50*	*0.09*	*-1.96*
8	*New(26)*	*X INV*	*0.81*	*0.50*	*0.79*	*0.90*	*1.00*	*0.89*	*0.61*	*0.82*	*1.00*	*0.21*	*0.39*	*0.18*	*-1.96*
9	66	X INV+Q	0.75	0.40	0.74	1.00	0.78	0.89	0.39	0.86	1.00	0.47	0.61	0.14	-1.55
10	*New(65)*	*X INV+Q*	*0.65*	*0.20*	*0.58*	*0.90*	*0.78*	*0.89*	*0.28*	*0.73*	*0.94*	*0.45*	*0.66*	*0.21*	*-0.85*
11	72	Relative clause	0.40	0.00	0.21	0.60	0.78	0.67	0.06	0.41	0.76	0.35	0.71	0.36	0.57
12	75	Relative clause	0.49	0.00	0.32	0.80	0.89	0.67	0.11	0.45	0.94	0.34	0.83	0.49	0.07

Table 9

Revised Grammar Ability Finder GP2: Form 2

Item #	Item #	Item type	IF	IF	IF	IF	IF	IF	IF	IF	IF	ID	ID	ID	Facets
(new)			total	partip	partip	partip	partip	partip	lower	mid	upper	mid-low	hi-low	hi-mid	item
				group 1	group 2	group 3	group 4	group 5	1/3	1/3	1/3				measure
1	36	Y-N INV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18
2	40	Y-N INV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18
3	54	Do-aux 2nd	0.89	0.80	0.84	0.90	1.00	1.00	0.78	0.91	1.00	0.13	0.22	0.09	-2.85
4	*New(51)*	*Do-aux 2nd*	*0.86*	*0.70*	*0.84*	*0.90*	*0.89*	*1.00*	*0.72*	*0.91*	*0.94*	*0.19*	*0.22*	*0.03*	*-2.45*
5	22	CAN+N	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18
6	*New(15)*	*CAN+N*	*0.88*	*0.50*	*0.89*	*1.00*	*1.00*	*1.00*	*0.67*	*0.95*	*1.00*	*0.29*	*0.33*	*0.05*	*-2.64*
7	*New(24)*	*X INV*	*0.81*	*0.50*	*0.79*	*1.00*	*1.00*	*0.78*	*0.50*	*0.91*	*1.00*	*0.41*	*0.50*	*0.09*	*-1.96*
8	26	X INV	0.81	0.50	0.79	0.90	1.00	0.89	0.61	0.82	1.00	0.21	0.39	0.18	-1.96
9	67	X INV+Q	0.72	0.30	0.68	1.00	0.78	0.89	0.33	0.82	1.00	0.48	0.67	0.18	-1.31
10	*New(65)*	*X INV+Q*	*0.65*	*0.20*	*0.58*	*0.90*	*0.78*	*0.89*	*0.28*	*0.73*	*0.94*	*0.45*	*0.66*	*0.21*	*-0.85*
11	*New(70)*	*Relative clause*	*0.40*	*0.00*	*0.26*	*0.60*	*0.67*	*0.67*	*0.06*	*0.36*	*0.82*	*0.31*	*0.77*	*0.46*	*0.57*
12	74	Relative clause	0.56	0.00	0.32	0.90	0.89	1.00	0.11	0.59	1.00	0.48	0.89	0.41	-0.33

Table 10

Revised Grammar Ability Finder GP2: Form 3, Business English

Item # (new)	Item #	Item type	IF total	IF partip group 1	IF partip group 2	IF partip group 3	IF partip group 4	IF partip group 5	IF lower 1/3	IF mid 1/3	IF upper 1/3	ID mid-low	ID hi-low	ID hi-mid	Facets item measure
1	37	Y-N INV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18
2	42	Y-N INV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18
3	*New(51)*	*Do-aux 2nd*	*0.86*	*0.70*	*0.84*	*0.90*	*0.89*	*1.00*	*0.72*	*0.91*	*0.94*	*0.19*	*0.22*	*0.03*	*-2.45*
4	*New(54)*	*Do-aux 2nd*	*0.89*	*0.80*	*0.84*	*0.90*	*1.00*	*1.00*	*0.78*	*0.91*	*1.00*	*0.13*	*0.22*	*0.09*	*-2.85*
5	17	CAN+N	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	-6.18
6	*New(15)*	*CAN+N*	*0.88*	*0.50*	*0.89*	*1.00*	*1.00*	*1.00*	*0.67*	*0.95*	*1.00*	*0.29*	*0.33*	*0.05*	*-2.64*
7	24	X INV	0.81	0.50	0.79	1.00	1.00	0.78	0.50	0.91	1.00	0.41	0.50	0.09	-1.96
8	*New(26)*	*X INV*	*0.81*	*0.50*	*0.79*	*0.90*	*1.00*	*0.89*	*0.61*	*0.82*	*1.00*	*0.21*	*0.39*	*0.18*	*-1.96*
9	*New(67)*	*X INV+Q*	*0.72*	*0.30*	*0.68*	*1.00*	*0.78*	*0.89*	*0.33*	*0.82*	*1.00*	*0.48*	*0.67*	*0.18*	*-1.31*
10	65	X INV+Q	0.65	0.20	0.58	0.90	0.78	0.89	0.28	0.73	0.94	0.45	0.66	0.21	-0.85
11	70	Relative clause	0.40	0.00	0.26	0.60	0.67	0.67	0.06	0.36	0.82	0.31	0.77	0.46	0.57
12	73	Relative clause	0.51	0.00	0.32	0.80	0.78	0.89	0.11	0.45	1.00	0.34	0.89	0.55	-0.03

Note that items 1-6 on each of the three forms all have very high IF values and virtually no discriminatory power (based on the pilot test investigation). These items were included in operational test forms for the following reasons. First, they represent the lowest word order rule stages tested in the current investigation. In operational test use, examinees who answer the more advanced items correctly (items 7-12) should also answer the majority, if not all, of these items correctly. If this is not the case, then serious consideration will need to be given to the further use of the test, because the theoretical basis for inferences will have been compromised. Second, these 'easier' items should serve the purpose of introducing examinees to the unique item response formats used on the test. Third, the use of only more difficult items might dissuade lower ability examinees from further completion of the web-based assessment battery, thereby defeating one test use objective (namely, to eventually link learners with learning resources). Finally, performance on these items may provide further indications about the English language abilities of the target population relative to the pilot participant sample.

Note also that a number of items from the pilot test (e.g., 30, 63, 68, 69) which had been retained for further consideration do not appear in the operational test forms. In sampling items for these forms, a decision had to be made as to which of the available items for a particular word order construct would best represent that construct, would discriminate the best among examinees at the appropriate levels of ability, and would elicit examinee performances in very similar ways compared with the same item on other forms of the test. For example, items 63, 68, and 69 all seemed to provide good discrimination between the lower and middle thirds of test scorers, and they were therefore retained as possible items representing the cancel-inversion word order rule. However, these three items also all showed the highest item facilities among the X INV+Q items, with even many of the lowest scoring examinees answering them correctly. For operational purposes, these three items would therefore prove less effective for distinguishing among the three ability level groups than would more difficult items (i.e., for this word order rule, items 65, 66, and 67). Therefore, in order to operationalize two items for this word order rule on each of the three test forms, items 65, 66, and 67 were utilized, and three new items were created, two of which replicated item 65 (the most difficult item) and one of which replicated item 67 (since items 66 and 67 showed very similar attributes). Similar decisions were made for each of the word order rule constructs, and items were sampled or revised accordingly.

Table 11 shows descriptive statistics and correlation coefficients comparing each of the proposed new test forms. These calculations are based on pilot test examinee performances for those items which were included intact in the three new forms, as well as on hypothesized performances for the new or revised items. These hypothesized performances were in turn estimated on the basis of examinee performances on the model item; as such, care should be taken in interpreting the statistics in Table 11, since actual performance patterns for each item may diverge from those predicted here. Nevertheless, these estimates do provide a point of reference for evaluating the likely usefulness and comparability of the three test forms.

In light of careful item sampling/revision, overall test performance outcomes are obviously predicted to be very similar, with virtually identical means and standard deviations. Despite the inclusion on each form of three non-discriminating items (i.e., those which did not discriminate among the pilot sample of examinees), Cronbach alpha reliability estimates indicate that the three forms would likely separate examinees with some consistency and to equivalent degrees. Pearson correlation coefficients between total scores on each of the forms and participant levels and “Mike” cloze scores reflect the same levels of moderate correlation as did the full-length test form. Further evidence for the parallel nature of the three forms can be seen in the correlation coefficients between total scores on each pair of test forms, which indicate that the three unique constellations of items are predicted to distinguish among examinees in very similar ways. This evidence suggests that the three forms would prove quite equivalent in placing examinees into Beginner, Intermediate, and Advanced levels. Of course, this interpretation also depends largely on the assumption that the several new items on each form would in fact perform in predicted ways.

Table 11

Descriptive Statistics and Pearson Correlations for Three New 12-item Forms

Statistic	Form 1	Form 2	Form 3
k	12	12	12
Mean	9.54	9.58	9.53
SD	2.19	2.10	2.12
Min	4.00	5.00	5.00
Max	12.00	12.00	12.00
Alpha	0.73	0.71	0.71
R (mike)	0.61	0.57	0.56
R (level)	0.63	0.71	0.68
R (Form 1)	1.00		
r (Form 2)	0.94	1.00	
r (Form 3)	0.92	0.98	1.00

In order to make the Beginner/Intermediate/Advanced distinctions on the basis of these three 12-item tests, it will also need to be determined at what cut-scores decisions will be made. Based on examinee performances in the current study, the following scoring bands are suggested as possibly reflecting the broad language ability differences of interest for operational test use: Beginner = 1 to 6 total points; Intermediate = 7 to 10 total points; Advanced = 11 to 12 total points. Thus, cut-scores would be set at 7 points and at 11 points. These cut-scores assume that the majority of examinees will be able to answer items 1-6 correctly (canonical, negation, and question formation items), fewer examinees will be able to additionally answer items 7-10 correctly (cancel inversion items), and fewer still will be able to additionally answer items 11 and 12 correctly (relative clause items). By way of example, based on these cut-scores and performances on the new test form #2, six examinees from the current investigation would have been placed as Beginners, all of whom were from participant level-grouping 1. Thirty examinees would have been placed as Intermediate learners, virtually all of whom were from participant level-groups 2, 3, and 4. Twenty-one examinees would have been placed as Advanced learners, all but two of whom were from participant level-groups 3, 4, and

5. Given the overall ability differences noted between level-groups 1 and 2 and the top three level-groups, as well as the similarities in average group ability estimates and variation in individual abilities among level-groups 3/4/5, these results seem justifiable. Moreover, if target examinees indeed reflect a range of lower proficiency learners in addition to those reflected in the current study, then these cut-scores would seem to make sense.

Table 12 shows decision dependability and error estimates (Subkoviak, 1988; see also discussion in Brown, 1996) associated with these two recommended cut-scores for each of the three forms. It should be kept in mind that these estimates are based on hypothesized performances for examinees on a number of items, and as such should not be interpreted as indicative of the scoring consistency which will be observed for operational administrations of the tests. Furthermore, reliability indices are always administration- and population-dependent and should therefore be calculated anew for future operational test use with the target learner population. Nevertheless, the estimates in Table 12 do give some indication of the extent to which the tests and associated cut-scores may inform decisions in a consistent manner. Note in Table 12 that a cut-score of 7 resulted in high dependability estimates (which may range from a low of 0.00 to a high of 1.00) and lower error estimates (which reflect a range around an examinee's score within which the examinee would likely score on a subsequent administration of the test) with the pilot examinee sample, while a cut-score of 11 showed substantially lower dependability and more extensive error. These findings reflect the fact that fewer items are contributing to the distinction between Advanced and Intermediate learners, and as a result, this decision will be less stable than the distinction between Beginner and Intermediate learners. However, both cut-scores do seem to be functioning within respectable levels of consistency, especially in light of the very small number of test items. Care should be taken to monitor examinees who score within (plus or minus) one point of each cut-score (and especially a cut-score of 11) during operational test use, in order to determine whether or not resulting placements are appropriate.

Table 12

Dependability and Error Estimates for Recommended Cut-Scores

	Cut-score = 7		Cut-score = 11	
Form	Dependability	*SEM*	Dependability	*SEM*
I	0.91	0.66	0.78	1.03
II	0.92	0.59	0.77	1.01
III	0.91	0.64	0.78	0.99

Finally, Table 13 shows that the three forms compiled within this approach to item sampling would apparently distinguish among examinees at each of the participant level-groupings in similar ways. Learners from the lowest pilot groups would be placed on average into the Beginner category, while learners from level-group 2 would be placed on average into the Intermediate category on all three test forms. Slightly greater variability in average level-group performance can be seen across the three forms at the upper three pilot ability levels (reflecting again the lower reliability of decisions at this cut-score), although the tests generally seem to be placing examinees on average into the Advanced category.

Table 13

Mean Scores by Participant Level-Grouping for Three New Test Forms

Group	Form 1	Form 2	Form 3
(1)	6.50	6.50	6.50
(2)	9.05	9.00	9.00
(3)	11.10	11.10	11.00
(4)	11.11	11.00	10.88
(5)	10.67	11.11	11.00

Sampling approach #2. A second approach to compiling three operational test forms might attempt to increase the reliability of distinctions between Intermediate and Advanced learners while maintaining the higher reliability of distinctions between

Beginner and Intermediate learners (as observed above). In order to do so, it is likely that increasing the number of relative clause items (those which seemed to discriminate the best between the highest scorers and all others on the pilot test) and decreasing the number of low- or non-discriminating items would result in overall higher levels of reliability as well as increased dependability of decisions made at the Intermediate/Advanced cut-score. While full operational test forms are not provided in the current report for these revisions, compilation of the new forms would proceed in a relatively straightforward manner.

First, items #1 and #5 could be removed from each of the three forms above, thereby eliminating one of the two Y-N INV items and one of the two CAN + N items (each of which showed zero discriminatory power for the current examinee sample). Remaining items for these word order rules would likely still prove sufficient for the purposes listed above (e.g., introducing examinees to the item response formats), although the test as a whole would become more difficult for most examinees. Next, two new relative clause items would need to be developed for each of the three test forms (unfortunately, sufficient numbers of this item type were not pilot tested in the current investigation for the development of three independent four-item test sections), including: (a) one pronoun-deletion relative clause item; and (b) one subject-, object-, or adverbial-pronoun relative clause item. These new items would need to replicate as closely as possible the linguistic and communicative contexts for those relative clause items retained in the three test forms above, in order to optimize the likelihood that target examinees would perform in predictable ways.

Table 14 displays the hypothesized descriptive statistics for these new test forms. These calculations are based on pilot test examinee performances on the subset of items on forms 1-3 above, with the important difference that items #1 and #5 have been removed and supplanted with hypothesized performances on two new relative clause items (as before, borrowing from the actual examinee performances on model items, in this case, existing items #11 and #12 on each form). Table 14 shows that the mean test scores would likely decrease (due to the inclusion of more difficult items), while variability in performances would increase (higher standard deviations due to the removal of non-discriminating items and greater variability in performances on the relative clause items). As predicted, the overall reliability of test scores would also increase substantially (Cronbach alpha for all forms = 0.81).

Table 14

Descriptive Statistics for Three New 12-Item Form

Statistic	Form 1	Form 2	Form 3
k	12	12	12
mean	8.56	8.54	8.44
SD	2.78	2.82	2.85
min	2.00	3.00	3.00
max	12.00	12.00	12.00
alpha	0.81	0.81	0.81

Naturally, cut-scores would also need to be adjusted for these new forms, in order to account for revised performance predictions associated with each category of target learners. Thus, learners would now be expected to perform the best on items 1-4, decreasingly well on items 5-8, and the worst on items 9-12. As such, new cut-scores would need to be set at 5, for the distinction between Beginner and Intermediate learners, and at 9, for the distinction between Intermediate and Advanced learners. Table 15 shows hypothesized decision dependability and error estimates associated with these new cut-scores. Note that dependability estimates increase for both cut-scores, as predicted, although the differences in comparison with dependability estimates for the previous operational forms (Table 12 above) are minimal. However, the error associated with a cut-score of 9 increases as well, suggesting that Intermediate/Advanced placement decisions would be slightly less stable than on the preceding forms.

Table 15

Dependability and Error Estimates for Recommended Cut-Scores

	Cut-score = 5		Cut-score = 9	
Form	Dependability	SEM	Dependability	SEM

I	0.94	0.68	0.80	1.24
II	0.93	0.75	0.80	1.26
III	0.93	0.75	0.80	1.27

CONCLUSION AND RECOMMENDATIONS FOR FURTHER RESEARCH AND TEST USE

As stated previously, which of the preceding approaches (or other possibilities) for item sampling/revision is adopted for the creation of operational test forms will depend on characteristics of the target learner population. Of course, it should also be reiterated that the calculations above were all based on observations from the current pilot test investigation, as well as on hypothesized performance patterns for a number of new items. The extent to which these predictions will be borne out can only be determined by further investigation, which is strongly encouraged prior to operational test use. However, based on findings from the current investigation, either of the two approaches outlined here would seem to result in test forms which could be easily operationalized in an on-line testing context and efficiently administered in a short amount of time, and which would satisfactorily distinguish among three broad English learner ability groups.

Several issues remain unresolved or in question for the grammar section of the *Ability Level Finder*, and these will need to be addressed in subsequent research. One major issue has to do with the characteristics of the target population and the extent to which they are related to those from the pilot test participants investigated in the current study. Future research should investigate the performances of examinees from a range of ability levels which are known to reflect those of the target population, and a variety of L1 background examinees will need to be included in each of these levels in order to investigate whether the current results were in some way biased by the limitations of available pilot test participants, especially in light of the fact that the overwhelming majority of pilot-test participants came from Korean and Japanese L1 backgrounds. Of course, this kind of research can only be undertaken where there is a target demographic basis for sampling participants.

Care will also need to be taken in operationalizing items from the current study in computer-mediated formats. The presentation of item prompts and target response words will need to remain clear, and where additional context is added (e.g., pictures), care must be taken to maintain the original communicative intent of the item. If target-like accuracy scoring is to be used, then the exact word order for target item responses will need to be provided for the computer's scoring algorithm. It should also be recalled that item instructions were carefully delivered by the test proctor in administering the pilot test GP1. For on-line delivery, test instructions will not only need to include additional instructions about the computer medium itself (e.g., on using the mouse to drag-and-drop), but they will also need to be very carefully formulated so that they may be understood by all examinees and for all item types. If feasible, this may call for instructions to be delivered in a variety of L1s. Of course, for any additional instructions that are developed, it will be necessary to make sure that the item construct is not being divulged, as this would endanger the validity of already fragile interpretations being made on the basis of only twelve test items.

Most importantly, the actual use of the tests and items for decision making purposes should be the focus of on-going validity inquiry (see Norris, 2000; 2004). Such inquiry would minimally address: (a) the extent to which examinees are responding in intended ways to each of the constructs represented in test items; (b) the extent to which the full range of target learners is performing in systematic and predicted ways with respect to the full set of test items and the two cut-scores; (c) the relationship between performance on the grammar test and all other components of the assessment battery; and (d) the extent to which eventual decisions that are made about examinees are appropriate and useful, both from the perspective of the test developers/users as well as from the point of view of the learners themselves.

REFERENCES

Brown, J. D. (1993). What are the characteristics of natural cloze tests? *Language Testing*, *10*(2), 93-116.

Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall.

Chapelle, C., Jamieson, J., & Hegelheimer, V. (2003). Validation of a web-based ESL test. *Language Testing*, *20*(4), 409-439.

Clahsen, H. (1984). The acquisition of German word order: A test case for cognitive approaches to second language acquisition. In R. Anderson (Ed.), *Second languages*. (pp. 219-242). Rowley, MA: Newbury House.

Clahsen, H., Meisel, J., & Pienemann, M. (1983). *Deutsch als Zweitsprache. Der Spracherwerb ausländischer Arbeiter*. Tübingen: Narr.

Ellis, R. (1989).Are classroom and naturalistic acquisition the same? A study of the classroom acquisition of German word order rules. *Studies in Second Language Acquisition*, *11*, 305-328.

Hudson, T. (1993). Nothing does not equal zero: Problems with applying developmental sequence findings to assessment and pedagogy. *Studies in Second Language Acquisition*, *15*, 461-493.

Ingold, C. (2002). The LangNet "Reading to the Four" project: Applied technology at higher levels of language learning. In B. Leaver & B. Shekhtman (Eds.), *Developing professional-level language proficiency* (pp. 141-155). Cambridge: Cambridge University Press.

Johnston, M. (1985). *Syntactic and morphological progressions in learner English*. Canberra: Department of Immigration and Ethnic Affairs.

Kanagy, R. (1994). Developmental sequences in learning Japanese: A look at negation. *Issues in Applied Linguistics*, *5*, 255-277.

Linacre, J. M. (1998). *Facets 3.17*. Computer program. Chicago: MESA Press.

Long, M. H., & Crookes, G. (1993). Units of analysis in syllabus design—The case for task. In G. Crookes & S. M. Gass (Eds.), *Tasks in a pedagogical context: Integrating theory and practice* (pp. 9-54). Clevedon, UK: Multilingual Matters.

McNamara, T. (1996). *Measuring second language performance*. New York: Longman.

Meisel, J., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in second language acquisition. *Studies in Second Language Acquisition*, *3*, 109-135.

Mellow, J. D. (1996). On the primacy of theory in applied studies: A critique of Pienemann and Johnston (1987). *Second Language Research*, *12*, 304-318.

Norris, J. M. (1996). *A validation study of the ACTFL Guidelines and the German Speaking Test*. Honolulu, HI: University of Hawaii, unpublished M.A. thesis.

Norris, J. M. (2001). Concerns with computer-adaptive oral proficiency assessment. *Language Learning & Technology*, *5*(2), 99-105. Available at: http://llt.msu.edu/vol5num2/norris/default.html/

Norris, J. M. (2000). Purposeful language assessment. *English Teaching Forum*, *38*(1), 18-23.

Norris, J. M. (2004). *Validity evaluation in foreign language assessment*. Honolulu, HI: University of Hawaii, unpublished doctoral dissertation.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417-528.

Norris, J. M., & Ortega, L. (2003). Defining and measuring SLA. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 717-761). Cambridge: Blackwell.

Pienemann, M. (1987). Determining the influence of instruction on L2 speech processing. *Australian Review of Applied Linguistics*, *9*, 92-122.

Pienemann, M. (1985). Learnability and syllabus construction. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 23-75). Clevedon, England: Multilingual Matters.

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics*, *10*, 217-244.

Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Philadelphia: John Benjamins.

Pienemann, M., & Johnston, M. (1987). Factors affecting the development of language proficiency. In D. Nunan (Ed.), *Applying second language acquisition research* (pp. 45-141). Adelaide: National Curriculum Resource Centre.

Pienemann, M., Johnston, M., & Brindley, G. (1988). Constructing an acquisition-based procedure for second language assessment. *Studies in Second Language Acquisition*, *10*, 217-243.

Pienemann, M., & Mackey, A. (1993). An empirical study of children's ESL development. In P. McKay (Ed.), *ESL development: Language and literacy in schools. Volume 2: Documents on bandscale development and language acquisition* (pp. 115-259). Canberra, Australia: National Languages & Literacy Institute of Australia and Commonwealth of Australia.

Roever, C. (2001). Web-based language testing. *Language Learning & Technology*, *5*(2), 84-94. Available at: http://llt.msu.edu/vol5num2/pdf/roever.pdf.

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

Spada, N., & Lightbown, P. (1999). Instruction, first language influence, and developmental readiness in second language acquisition. *The Modern Language Journal*, *83*, 1-22.

Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement*, *25*, 47-55.

APPENDIX A

TEST YOUR ENGLISH

Ability Finder GP1

John M. Norris

CONSENT FORM

Please read the following statement. Then sign and date this form.

Thank you for volunteering to participate in this project. Your answers on the "Test Your English" ability finder test will be used for research purposes. Your name will be used for our identification purposes only. Except for the researchers, no one will see individual names or scores from this test. If you have questions at any time during the test, please ask the test administrator.

"I have read and understood the statement above, and I agree to the use of my answers on this test for research purposes."

______________________________	______________
(volunteer's signature)	**(date)**

TEST YOUR ENGLISH

Ability Finder GP1

Instructions: This is an English grammar test. There are 7 sections. The test administrator will explain the instructions for each section before you start. After you finish each section, wait until the test administrator tells you to go on to the next section. Do not go back and work on earlier sections of the test. If you do not understand any of the instructions, or if you have any questions, please ask the test administrator for help. Try to do your best on all of the items on the test. The test will take about 45 minutes.

Background information: Before beginning the test, please answer the following questions.

1. What is your name? ____

2. What is your age? _______________ years

3. Where are you taking classes?

_________ HELP _________ ELI

4. What classes are you taking at HELP or the ELI this semester?
(Please list the course number for each class; for example: *HELP 50*, *ELI 82*)

5. What is your native (1st) language? ______________________________

6. How many years have you studied English? _______________ years

7. What was your best overall score on the TOEFL?

__________ points __________(have not taken)

STOP! DO NOT TURN THE PAGE UNTIL ASKED TO DO SO.

Section I. Fill in the blanks.

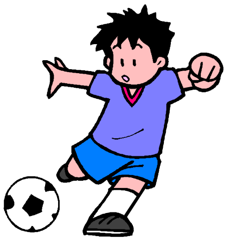
Instructions: Look at the picture for each item. Then fill in the blanks to make a good sentence about the picture. Write only one word in each blank. Please write clearly and spell correctly.

Example item:

Those _____ _____ hungry.

Answer: Those _fish_ _are_ hungry.

Item 1:



This ___________ ___________ soccer every Saturday.

Item 2:

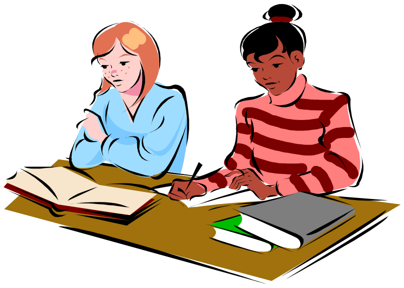
Monday

Tuesday

Wednesday

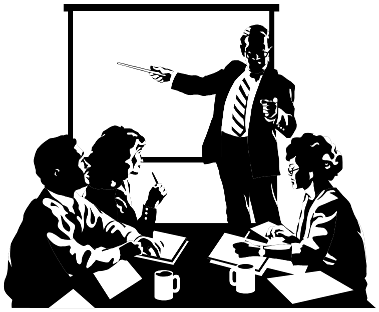
That ___________ ___________ to people on the phone every day.

Item 3:



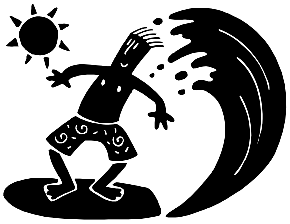
Those ____________ ____________ a lot!

Item 4:



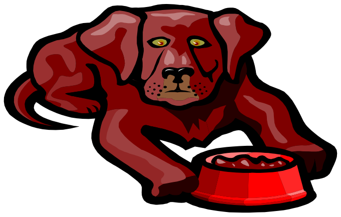
Some ____________ ____________ coffee while they work.

Item 5:



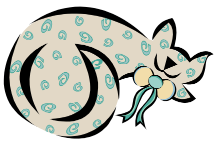
That ____________ ____________ to surf!

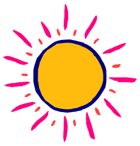
Item 6:



This ____________ ____________ too much.

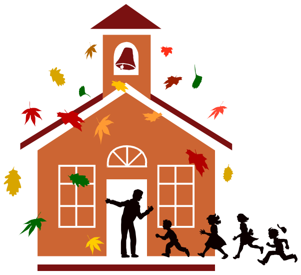
Item 7:





Many ____________ ____________ during the day.

<u>Item 8</u>:



Most _____________ _____________ to school.

STOP! DO NOT TURN THE PAGE UNTIL ASKED TO DO SO.

Section II. Make a sentence.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to make an appropriate response in the blanks provided. Words in the same box stay together.

Example:

Classmate: "Amy is studying tonight."

Words:	a test	has	tomorrow	she

Friend: "__She__ __has__ __a test__ __tomorrow__."

Item 9:

Friend: "George, you might be late."

Words:	in	my flight	thirty minutes	leaves

George: "___________ ___________ ___________ ___________."

Item 10:

Mother: "Sandra gets good grades in school."

Words:	studies	a lot	she

Father: "___________ ___________ ___________."

Item 11:

Office manager: "We need a new office copier."

Words:	will	a copier	I	on the internet	order

Secretary: "___________ ___________ ___________ ___________ ___________."

Item 12:

Couple: “We want a reservation for 7:00 p.m.”

Words:	available	there	tables	no	are

Waiter: “___________ ___________ ___________ ___________ ___________.”

Item 13:

Coach: “That is the worst football team in the state.”

Words:	any	never	they	win	games

Assistant coach: “__________ __________ __________ __________ __________.”

Item 14:

Caller: “May I speak with your father?”

Words:	parents	are	at home	my	not

Daughter: “___________ ___________ ___________ ___________ ___________.”

Item 15:

Doctor: “So, tell me about your new diet.”

Words:	no	I	vegetables	eat

Patient: “___________ ___________ ___________ ___________.”

Item 16:

Friend 1: “Russell and Theresa are a strange couple.”

Words:	belong	don’t	together	they

Friend 2: “___________ ___________ ___________ ___________.”

Item 17:

Co-worker: “Mary, you can use the phone if you like.”

Words:	like	cell phones	not	do	I

Mary: “___________ ___________ ___________ ___________ ___________.”

Item 18:

Client: “That employer asked about my ethnicity.”

Words:	that question	have	you	don’t	to answer

Lawyer: “___________ ___________ ___________ ___________ ___________.”

Item 19:

Traveler: “There is a flight to Madrid, right?”

Words:	anymore	we	not	to Spain	fly	do

Ticket agent: “_________ _________ _________ _________ _________ _________.”

Item 20:

Grandfather: “Sally seems very lazy today.”

Words:	to	go	won’t	school	she

Grandmother: “__________ __________ __________ __________ __________.”

Item 21:

Sister: “Tyler has a bad tooth.”

Words:	eat	not	he	should	sweets

Mother: “___________ ___________ ___________ ___________ ___________.”

Item 22:

Co-worker 1: “My friends were disappointed with the meal last night.”

Words:	like	the restaurant	did	they	not

Co-worker 2: “__________ __________ __________ __________ __________.”

Item 23:

Teacher: “The boy will have trouble finishing the test.”

Words:	read	enough	fast	doesn’t	he

Teacher’s aid: “__________ __________ __________ __________ __________.”

STOP! DO NOT TURN THE PAGE UNTIL ASKED TO DO SO.

Section III. Complete the sentence.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to complete the sentence that is started for you. Words in the same box stay together.

Example:

Jogger 1: “My dog runs with me every day.”

Words:	would	my dog	with me	run

Jogger 2: “*I wish* __ my dog __ __ would __ __ run __ __ with me __.”

Item 24:

Secretary: “Tom has worked here for 40 years.”

Words:	is	old	he	how

Office clerk: “*I wonder* ____________ ____________ ____________ ____________.”

Item 25:

Travel agent: “The trip to Tokyo lasts seven hours.”

Words:	a movie	will	they	show

Traveler: “*I hope* ____________ ____________ ____________ ____________.”

Item 26:

Roommate: "Hey Andrew, your teacher called."

Words:	in school	not	why	been	have	I

Andrew:
"*He wants to know* ________ ________ ________ ________ ________ ________."

Item 27:

Student 1: "On Tuesday we have the final exam."

Words:	are	for it	prepared	you

Student 2: "*I know* ____________ ____________ ____________ ____________."

Item 28:

Veterinarian: "That dog looks very sad."

Words:	won't	why	play	he

Assistant: "*I wonder* ____________ ____________ ____________ ____________."

Item 29:

Coach: "The team is really angry."

Words:	to play	was	I	not	why	able

Player:
"*They want to know* ________ ________ ________ ________ ________ ________."

Item 30:

Office manager: "I heard Terri is sick today."

Words:

you	whether	the phones	answer	could

Boss: "*She asked* __________ __________ __________ __________ __________."

Item 31:

Accountant: "Any messages from the boss?"

Words:

were	whether	at	you	the interview

Assistant:
"*She wants to know* __________ __________ __________ __________ __________."

Item 32:

Flight attendant: "This is a short flight to New York."

Words:

will	a meal	they	whether	serve

Traveler: "*I wonder* _________ __________ __________ __________ __________."

STOP! DO NOT TURN THE PAGE UNTIL ASKED TO DO SO.

Section IV. Ask a question.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to ask a good question in response to the sentence. Words in the same box stay together.

Example:

Waitress: "There are no free tables inside."

Words:

a table	there	outside	is

Customer: "__Is__ __there__ __a table__ __outside__ **?**"

Item 33:

Girl: "I like to go to the beach."

Words:

volleyball	you	play	do

Boy: "__________ __________ __________ __________ **?**"

Item 34:

Father: "I'm sorry, David and Susan are not at home."

Words:

they	did	go	the museum	to

Caller: "__________ __________ __________ __________ __________ **?**"

Item 35:

Teacher: "I think Maria will succeed in college."

Words:

good student	she	a	is

College recruiter: "__________ __________ __________ __________ **?**"

Item 36:

Student: “Thanks for the answer, Professor Jones.”

Words: | other questions | there | are | any |

Professor Jones: “___________ ___________ ___________ ___________ **?**”

Item 37:

Friend 1: “Amanda never has any money.”

Words: | she | work | doesn’t |

Friend 2: “___________ ___________ ___________ **?**”

Item 38:

Secretary: “You have a meeting tomorrow at 11:00 a.m.”

Words: | by noon | end | will | the meeting |

Boss: “___________ ___________ ___________ ___________ **?**”

Item 39:

Professor: “Graham speaks German very well.”

Words: | he | been | has | to Germany |

Student: “___________ ___________ ___________ ___________ **?**”

Item 40:

Daughter: "I'm not hungry for lunch yet."

Words:

a	you	did	big breakfast	eat

Father: "__________ __________ __________ __________ __________ **?**"

Item 41:

Brother: "Carmen is not at the station."

Words:

a different bus	have	she	taken	could

Mother: "__________ __________ __________ __________ __________ **?**"

Item 42:

Employer: "You are perfect for the job!"

Words:

or	in August	the job	start	September	does

Job applicant:
"__________ __________ __________ __________ __________ __________**?**"

Item 43:

Doctor: "Your uncle was very sick."

Words:

should	I	called	have	an ambulance

Nephew: "__________ __________ __________ __________ __________ **?**"

Item 44:

Employee: “Everyone at work is worried about the flu.”

Words:	had	you	a vaccination	have

Nurse: “___________ ___________ ___________ ___________ **?**”

Item 45:

Ticket agent: “There are only two more flights to New York this evening.”

Words:	to New York	is	back	the last flight	when

Traveler: “___________ ___________ ___________ ___________ ___________ **?**”

Item 46:

Students: “We all got that question wrong on the test.”

Words:	the	knows	who	correct answer

Teacher: “___________ ___________ ___________ ___________ **?**”

Item 47:

Friend 1: “Erica and Doug are always smiling.”

Words:	are	why	so happy	they

Friend 2: “___________ ___________ ___________ ___________ **?**”

Item 48:

Worker: “I am really tired of working.”

Words: a vacation | don’t | take | why | you

Boss: “___________ ___________ ___________ ___________ ___________ **?**”

Item 49:

Son: “Finally we are on summer vacation!”

Words: will | when | you | back to school | go

Father: “___________ ___________ ___________ ___________ ___________ **?**”

Item 50:

Daughter: “My mother was named after her grandmother.”

Words: she | how | her name | spell | does

Friend: “___________ ___________ ___________ ___________ ___________ **?**”

Item 51:

Roommate 1: “We are out of cat food.”

Words: can | eat | the cat | what

Roommate 2: “___________ ___________ ___________ ___________ **?**”

Item 52:

Friend 1: “I will buy all of the decorations and drinks.”

Words:

who	to the party	we	invite	should

Friend 2: “___________ ___________ ___________ ___________ ___________ **?**”

Item 53:

Grandmother: “All of my grandchildren are home for the holidays.”

Words:

are	they	where	sleeping

Neighbor: “___________ ___________ ___________ ___________ **?**”

Item 54:

Professor 1: “I won’t be there on Saturday.”

Words:

you	why	attend	can’t	the lecture

Professor 2: “__________ __________ __________ __________ __________ **?**”

Item 55:

Student: “I cannot finish writing the paper.”

Words:

won’t	by Friday	able	be	who else	to finish

Teacher:
“__________ __________ __________ __________ __________ __________ **?**”

<u>Item 56</u>:

Surfer: “This is a really dangerous beach.”

Words:	seen	the sharks	where	were

Tourist: “___________ ___________ ___________ ___________ **?**”

<u>Item 57</u>:

Neighbor 1: “I think someone broke into our house.”

Words:	didn’t	the police	you	why	call

Neighbor 2: “__________ __________ __________ __________ __________ **?**”

<u>Item 58</u>:

Flight attendant: “We will start the second movie now.”

Words:	it	a long flight	is	isn’t	this

Traveler: “___________ ___________ ___________, ___________ ___________ **?**”

<u>Item 59</u>:

Friend 1: “I saw *Star Wars* again yesterday.”

Words:	a	was	not	that	great movie

Friend 2: “___________ ___________ ___________ ___________ ___________ **?**”

Item 60:

Student: “We had too much homework last night.”

Words: finish you your didn’t homework

Professor: “___________ ___________ ___________ ___________ ___________?”

Item 61:

Movie critic 1: “I thought Dustin Hoffman was great.”

Words: was great wasn’t he he

Movie critic 2: “__________ __________ __________, __________ __________ **?**”

Item 62:

Co-worker: “She is still feeling ill.”

Words: been not she has to the doctor

Boss: “___________ ___________ ___________ ___________ ___________ **?**”

STOP! DO NOT TURN THE PAGE UNTIL ASKED TO DO SO.

Section V. Complete the question.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to complete the question that is started for you. Words in the same box stay together.

Example:

Employee: "I need more space at work."

Words:	office	move	to	like	to another

Boss: "*Would you* like to move to another office **?**"

Item 63:

Mother: "My son is losing a lot of weight."

Words:	has	whether	sick	been	he

Nurse:
"*Do you know* __________ __________ __________ __________ __________ **?**"

Item 64:

Friend 1: "Miguel said that his parents are not at home."

Words:	they	where	went

Friend 2: "*Does he know* __________ __________ __________ **?**"

Item 65:

Interviewer: "You must have some questions about the job."

Words:	the job	much	pays	how

Job applicant: "*Can you tell me* __________ __________ __________ __________**?**"

Item 66:

Airline agent: "The flight left Singapore at 10:00 a.m."

Words:	arriving	it	when	is	in the U.S.

Caller:
"*Do you know* __________ __________ __________ __________ __________ **?**"

<u>Item 67</u>:

Student 1: "The teacher told me to bring my lunch for the trip tomorrow."

Words:	are	where	meeting	we

Student 2: "*Did she say* ___________ ___________ ___________ ___________ **?**"

<u>Item 68</u>:

Employee 1: "Let's order pizza for lunch."

Words:	wants	the boss	what

Employee 2: "*Should we ask* ___________ ___________ ___________ **?**"

<u>Item 69</u>:

Friend 1: "Max and Nina saw that movie yesterday."

Words:	good	was	whether	it

Friend 2:
"*Did they tell you* ___________ ___________ ___________ ___________ **?**"

STOP! DO NOT TURN THE PAGE UNTIL ASKED TO DO SO.

Section VI. Complete the sentence.

Instructions: Read the question for each item. Then look at the words in the boxes. Move the boxes to complete the sentence that is started for you. Words in the same box stay together. BE CAREFUL: There are more boxes than blanks. You will not be able to use all of the boxes, but you must fill all of the blanks.

Example:

Office manager: “Who will take calls today?”

Words:

on the phone	who	will

Sentence:
“*The man* on the phone will.”

Item 70:

Employee: “Who is that new woman working in the sales department?”

Words:

the boss	about	spoke	the woman	her	is

Receptionist:
“*She* ___________ ___________ ___________ ___________ ___________.”

Item 71:

Former student: “Where are you studying?”

Words:

that	the college	went	to	go	you	to

Student:
“*I* __________ __________ __________ __________ __________ __________.”

Item 72:

Clerk: “Can you describe the book?”

Words:

is	are looking	old	for	we	it

Customer:
“*The book* ___________ ___________ ___________ ___________ ___________.”

Item 73:

Client: “Is this a successful company?”

Words:	owns	a lot	she	earns	who	the business

Stock broker:
“*Well, the woman* __________ __________ __________ __________ ___________.”

<u>Item 74</u>:

Friend 1: “How do you save money on groceries?”

Words:	is	it	we	shop	where	cheap

Friend 2:
“*The store* ___________ ___________ ___________ ___________ ___________.”

<u>Item 75</u>:

Mother: “Are any of these players really good?”

Words:	is	sister	great	he	whose	plays soccer

Coach:
“*The boy* ___________ ___________ ___________ ___________ ___________.”

STOP! DO NOT TURN THE PAGE UNTIL ASKED TO DO SO.

Section VII. Fill in the blanks.

Instructions: Read the following two paragraphs. Then write a single word in each blank. Try to choose words that make the most sense in each sentence.

"Mike"

Mike worked very hard the first week on the job, harder than I did. But after a month had passed, his attitude toward the job had soured, and his disposition with it. He found the work we were asked to do boring ____________ monotonous. One day he told me that he burned his finger ____________ he wasn't paying attention to his work. He kept saying ____________ couldn't stand being cooped up, that he wasn't suited to the ____________. As the weeks of the summer went by, he became more and ____________ short-tempered, more anxious to get out. One day in sheer frustration, ____________ asked Tom, an older worker, "Do you think I should take the grade coil wrapper job in Lawrence?" Tom said, "No." Then Mike asked, "How do you think I can get ahead?" Tom answered, "Quit, go to school." "I can't, I need the money," Mike answered, discouraged. Tom shrugged his shoulders ____________ said, as he walked away, "Well, then you'll just have to ____________ on doing what you're doing now."

Mike was still working when ____________ left Western Electric, but he was complaining about it more often. Three ____________ later, he quit and took a job delivering milk. He held the new job for only four more months because his temper got the best of him one day and got him fired. Poor Mike is unemployed now, but he is thinking of going back to Western for another try, if they will have him.

THIS IS THE END OF THE TEST. THANK YOU FOR YOUR HELP!

APPENDIX B

GP1 ADMINISTRATION INSTRUCTIONS

TEST YOUR ENGLISH: ABILITY FINDER GP1

ADMINISTRATION INSTRUCTIONS

Materials needed:

__________ Sufficient copies of the test, stapled in the upper left-hand corner (Don't copy front to back—single-sided only)

__________ Pencils for student use

__________ Receipt forms for collecting student signatures and dates

__________ Sufficiently large stack of ten-dollar bills

1. Seat students so that they cannot see any other students' test forms (also, no dictionaries, etc. allowed or necessary)

2. After all students have arrived, distribute test forms and pencils—tell students to wait for further instructions.

3. When all students have the test form, ask them to open the test to page 2. Read the Consent Form aloud while students read along with you. Ask if there are any questions about the form, then make sure that they all sign and date the form.

4. Ask students to turn to page 3. Go over the general test instructions aloud while they read along. Make sure the students understand that you will tell them when they are to go on to the next section, and that they should wait until you do so. Make sure they also understand that they should not go back to any previous sections (this is important, since later sections may give clues for answering earlier sections). Ask if there are any questions before you begin the test. Answer questions and then have students fill in the demographic data section on page 3.

5. Once all students have completed page 3, ask them to turn to page 4. Read through the instructions and look at the example together. Make sure they understand that they have to write one and only one word in each of the two blanks for each item. Tell students to write legibly and spell correctly. Make sure that students understand that they are making a sentence that makes sense as a response to the first sentence. Do not explain anything about how students should select the words, etc. (this is what we are testing). Ask if there are any questions. Tell students to begin. Monitor students as they work on the test, making sure that none try to move ahead to the next section.

6. Once all students are finished with the 8 items in section I, ask them to turn to page 7. Read through the instructions and go over the example together. Make sure students understand that words which appear in the same box must stay together in a single blank in their answers (use the example to show what you mean). Be careful not to explain the construct here or in subsequent sections. Tell students to begin. Monitor as above.

7. Once all students have finished section II, ask them to turn to page 11. Go over the instructions and the example. Crucial here is that students understand that the sentence is already started for them. Use the example to make sure that they see the first few words are provided and that the answer is a continuation of these words. Point out the sentence-beginning words on items 24 and 25 as well. Tell students to begin. Continue monitoring.

8. Once all students have finished section III, ask them to turn to page 14. Go over instructions and the example. It is very important that you emphasize here that students are now asking questions. Point this out (with rising intonation, etc.) in the example, and indicate the question marks at the end of each series of blanks so that students all obviously understand that they are doing something different than before. Don't say anything about how questions should be formed (or answer any substantive questions about the same). Tell students to begin and continue monitoring. Make sure students are not going back to previous sections.

9. Once all students have finished section IV, ask them to turn to page 22. Go over the instructions and example, making sure again that students understand that the questions in this section are started for them with a few words. Point this out on the first several items. Also, if any students try to go back to previous sections, I reckon it will be here, since they now see question formation given in the first few words—monitor accordingly. Tell students to begin.

10. Once all students have finished section V, ask them to turn to page 24. Go over instructions and the example. Make sure they notice the differences in this section: (1) that the sentences are started for them, (2) that there are now more boxes than there are blanks available (so they have to pick only the right boxes). Reiterate here that students should keep words that appear in the same box together in a single blank. Tell students to begin.

11. Once all students have finished section VI, ask them to turn to page 26. Go over instructions. Stress to students that they should read through the entire passage first, then go back and try to fill in the blanks. Make sure they understand that they can only put one word per blank. Tell students that when they finish individually they can come and get paid by you and leave. Tell students to begin. As students finish, collect their tests and then make sure that they sign and date the receipt (and that they get paid). Thanks students for their participation.

APPENDIX C

TEST YOUR ENGLISH

ABILITY FINDER GP2

(REVISED 12-ITEM TEST FORMS 1, 2, 3)

TEST YOUR ENGLISH

Ability Finder GP2

General Instructions: This is an English grammar test. There are 5 sections. Before you begin each section, carefully read the instructions for that section. Try to do your best on all 12 of the items on the test.

Section I. Ask a question.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to ask a good question in response to the sentence. Words in the same box stay together.

Example:

Mother: “There is a lot of food in the refrigerator.”

Words:	the store	who	to	went

Father: “___Who___ ___went___ ___to___ ___the store___ **?**”

Item 1:

Teacher: “I think Maria will succeed in college.”

Words:	good student	she	a	is

College recruiter: “___________ ___________ ___________ ___________ **?**”

Item 2:

Girl: “I like to go to the beach.”

Words:	volleyball	you	play	do

Boy: “___________ ___________ ___________ ___________ **?**”

Item 3:

Friend 1: "I will buy all of the decorations and drinks."

Words:	who	to the party	we	invite	should

Friend 2: "___________ ___________ ___________ ___________ ___________ **?**"

Item 4:

Roommate 1: "We are out of cat food."

Words:	can	eat	the cat	what

Roommate 2: "___________ ___________ ___________ ___________ **?**"

Section II. Make a sentence.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to make an appropriate response in the blanks provided. Words in the same box stay together.

Example:

Classmate: “Amy is studying tonight.”

Words:	a test	has	tomorrow	she

Friend: “ She has a test tomorrow .”

Item 5:

Sister: “Tyler has a bad tooth.”

Words:	eat	not	he	should	sweets

Mother: “___________ ___________ ___________ ___________ ___________.”

Item 6:

Doctor: “So, tell me about your new diet.”

Words:	no	I	vegetables	eat

Patient: “___________ ___________ ___________ ___________.”

Section III. Complete the sentence.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to complete the sentence that is started for you. Words in the same box stay together.

Example:

Jogger 1: “My dog runs with me every day.”

Words:	would	my dog	with me	run

Jogger 2: “*I wish* my dog would run with me.”

Item 7:

Wife: “We should drive to New York City.”

Words:	far	is	how	it

Husband: “*I wonder* ____________ ____________ ____________ ____________.”

Item 8:

Roommate 1: “We got a letter from the rental agency.”

Words:	the check	we	why	sent	not	have

Roommate 2:
“*They want to know* ________ ________ ________ ________ ________ ________.”

Section IV. Complete the question.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to complete the question that is started for you. Words in the same box stay together.

Example:

Employee: "I need more space at work."

Words:	office	move	to	like	to another

Boss: "*Would you* like to move to another office **?**"

Item 9:

Airline agent: "The flight left Singapore at 10:00 a.m."

Words:	arriving	it	when	is	in the U.S.

Caller:
"*Do you know* __________ __________ __________ __________ __________ **?**"

Item 10:

Art dealer: "That's a very beautiful painting."

Words:	the painting	much	how	costs

Customer:
"*Can you let me know* __________ __________ __________ __________ **?**"

Section V. Complete the sentence.

Instructions: Read the question for each item. Then look at the words in the boxes. Move the boxes to complete the sentence that is started for you. Words in the same box stay together. **BE CAREFUL**: There are more boxes than blanks. You will not be able to use all of the boxes, but you must fill all of the blanks.

Example:

Office manager: “Who will take calls today?”

Words:

on the phone	who	will

Sentence:
“*The man* on the phone will.”

Item 11:

Clerk: “Can you describe the book?”

Words:

is	are looking	old	for	we	it

Customer:
“*The book* ___________ ___________ ___________ ___________ ___________.”

Item 12:

Mother: “Are any of these players really good?”

Words:

is	sister	great	he	whose	plays soccer

Coach:
“*The boy* ___________ ___________ ___________ ___________ ___________.”

TEST YOUR ENGLISH

Ability Finder GP2

FORM 2

John M. Norris

TEST YOUR ENGLISH

Ability Finder GP2

General Instructions: This is an English grammar test. There are 5 sections. Before you begin each section, carefully read the instructions for that section. Try to do your best on all 12 of the items on the test.

Section I. Ask a question.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to ask a good question in response to the sentence. Words in the same box stay together.

Example:

Mother: “There is a lot of food in the refrigerator.”

Words:

the store	who	to	went

Father: “__Who__ __went__ __to__ __the store__ **?**”

Item 1:

Student: “Thanks for the answer, Professor Jones.”

Words:

other questions	there	are	any

Professor Jones: “__________ __________ __________ __________ **?**”

Item 2:

Daughter: “I’m not hungry for lunch yet.”

Words:

a	you	did	big breakfast	eat

Father: “__________ __________ __________ __________ __________**?**”

Item 3:

Professor 1: "I won't be there on Saturday."

Words: | you | why | attend | can't | the lecture |

Professor 2: "__________ __________ __________ __________ __________ **?**"

Item 4:

Friend 1: "I hope this band plays good music."

Words: | play | what | the band | should |

Friend 2: "__________ __________ __________ __________ **?**"

Section II. Make a sentence.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to make an appropriate response in the blanks provided. Words in the same box stay together.

Example:

Classmate: “Amy is studying tonight.”

Words:	a test	has	tomorrow	she

Friend: “ She has a test tomorrow .”

Item 5:

Co-worker 1: “My friends were disappointed with the meal last night.”

Words:	like	the restaurant	did	they	not

Co-worker 2: “__________ __________ __________ __________ __________.”

Item 6:

Customer: “I would like a loaf of wheat bread, please.”

Words:	have	wheat bread	no	we

Baker: “__________ __________ __________ __________.”

Section III. Complete the sentence.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to complete the sentence that is started for you. Words in the same box stay together.

Example:

Jogger 1: “My dog runs with me every day.”

Words:

would	my dog	with me	run

Jogger 2: “*I wish* my dog would run with me.”

Item 7:

Tour guide: “Many of these office buildings were built in the 1950s.”

Words:

are	how	they	tall

Tourist: “*I wonder* ___________ ___________ ___________ ___________.”

Item 8:

Roommate: “Hey Andrew, your teacher called.”

Words:

in school	not	why	been	have	I

Andrew:
“*He wants to know* ________ ________ ________ ________ ________ ________.”

Section IV. Complete the question.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to complete the question that is started for you. Words in the same box stay together.

Example:

Employee: “I need more space at work.”

Words:	office	move	to	like	to another

Boss: “*Would you* like to move to another office **?**”

Item 9:

Student 1: “The teacher told me to bring my lunch for the trip tomorrow.”

Words:	are	where	meeting	we

Student 2: “*Did she say* ___________ ___________ ___________ ___________ **?**”

Item 10:

Friend 1: “Max and Nina saw that movie yesterday.”

Words:	the movie	how	lasts	long

Friend 2: “*Did they tell you* __________ __________ __________ __________ **?**”

Section V. Complete the sentence.

Instructions: Read the question for each item. Then look at the words in the boxes. Move the boxes to complete the sentence that is started for you. Words in the same box stay together. **BE CAREFUL**: There are more boxes than blanks. You will not be able to use all of the boxes, but you must fill all of the blanks.

Example:

Office manager: "Who will take calls today?"

Words:	on the phone	who	will

Sentence:
"*The man* __on the phone__ ___will___."

Item 11:

Student: "What are we supposed to do for tomorrow?"

Words:	yesterday	it	I	is	your homework	assigned

Teacher:
"*The chapter* ___________ ___________ ___________ ___________ ___________."

Item 12:

Friend 1: "How do you save money on groceries?"

Words:	is	it	we	shop	where	cheap

Friend 2:
"*The store* ___________ ___________ ___________ ___________ ___________."

TEST YOUR ENGLISH

Ability Finder GP2

FORM 3, Business English

John M. Norris

TEST YOUR ENGLISH

Ability Finder GP2

General Instructions: This is an English grammar test. There are 5 sections. Before you begin each section, carefully read the instructions for that section. Try to do your best on all 12 of the items on the test.

Section I. Ask a question.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to ask a good question in response to the sentence. Words in the same box stay together.

Example:

Mother: “There is a lot of food in the refrigerator.”

Words:

the store	who	to	went

Father: “___Who___ ___went___ ___to___ ___the store___ **?**”

Item 1:

Friend 1: “Amanda never has any money.”

Words:

she	working	is

Friend 2: “___________ ___________ ___________ **?**”

Item 2:

Employer: “You are perfect for the job!”

Words:

or	in August	the job	start	September	does

Job applicant:
“___________ ___________ ___________ ___________ ___________ ___________ **?**”

Item 3:

Co-worker 1: “Andres has a lot of experience.”

Words:	doesn’t	apply	why	he	for the job

Co-worker 2: “__________ __________ __________ __________ __________ **?**”

Item 4:

Office manager: “We have to send a letter to the employees.”

Words:	say	the letter	should	what

Company executive: “___________ ___________ ___________ ___________ **?**”

Section II. Make a sentence.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to make an appropriate response in the blanks provided. Words in the same box stay together.

Example:

Classmate: “Amy is studying tonight.”

Words:	a test	has	tomorrow	she

Friend: “__She__ __has__ __a test__ __tomorrow__.”

Item 5:

Co-worker: “Mary, you can use the phone if you like.”

Words:	like	cell phones	not	do	I

Mary: “___________ ___________ ___________ ___________ ___________.”

Item 6:

New worker: “So, Jane, tell me about the year-end bonus at this job.”

Words:	no	get	bonuses	we

Jane: “___________ ___________ ___________ ___________.”

Section III. Complete the sentence.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to complete the sentence that is started for you. Words in the same box stay together.

Example:

Jogger 1: “My dog runs with me every day.”

Words:

would	my dog	with me	run

Jogger 2: “*I wish* my dog would run with me.”

Item 7:

Secretary: “Tom has worked here for 40 years.”

Words:

is	old	he	how

Office clerk: “*I wonder* ____________ ____________ ____________ ____________.”

Item 8:

Accountant: “Any messages from the boss?”

Words:

you	paid	have	why	the bills	not

Assistant:
“*She wants to know* ________ ________ ________ ________ ________ ________.”

Section IV. Complete the question.

Instructions: Read the first sentence for each item. Then look at the words in the boxes. Move the words to complete the question that is started for you. Words in the same box stay together.

Example:

Employee: “I need more space at work.”

Words:

office	move	to	like	to another

Boss: “*Would you* like to move to another office **?**”

Item 9:

Employee 1: “Let’s order pizza for lunch.”

Words:

is	the boss	when	eating

Employee 2:
“*Should we ask* ___________ ___________ ___________ ___________ **?**”

Item 10:

Interviewer: “You must have some questions about the job.”

Words:

the job	much	pays	how

Job applicant:
“*Can you tell me* ___________ ___________ ___________ ___________ **?**”

Section V. Complete the sentence.

Instructions: Read the question for each item. Then look at the words in the boxes. Move the boxes to complete the sentence that is started for you. Words in the same box stay together. **BE CAREFUL**: There are more boxes than blanks. You will not be able to use all of the boxes, but you must fill all of the blanks.

Example:

Office manager: “Who will take calls today?”

Words:	on the phone	who	will

Sentence:
“*The man* on the phone will.”

Item 70:

Employee: “Who is that new guy working in the sales department?”

Words:	the boss	about	spoke	the man	him	is

Receptionist:
“*He* __________ __________ __________ __________ __________.”

Item 12:

Client: “Is this a successful company?”

Words:	owns	a lot	she	earns	who	the business

Stock broker:
“*Well, the woman* __________ __________ __________ __________ __________.”