

## Written Discourse Completion Tasks in the Foreign Language Examination (YDS)

Zülal AYAR<sup>1</sup>

<sup>1</sup>Ph.D., Ankara University, Ankara, TURKEY  
ayarz@ankara.edu.tr

ORCID: 0000-0002-9029-7164

### Article information

Submission	19/11/2020	Revision received	08/02/2021
Acceptance	14/03/2020	Publication date	16/04/2021

**Keywords:**  
pragmatic  
assessment,  
international  
language test,  
foreign language  
examination, written  
discourse  
completion task,  
YDS

**Abstract:** As the most prestigious and popular standardized achievement test to certify examinees' proficiency of the English language at the national level, Foreign Language Examination (YDS) has been mostly taken by academic staff, undergraduate and graduate students, state employees, and military personnel for years in Turkey. The current study set out to address multiple-choice written discourse completion tasks (WDCTs) in YDS from 2015 to 2019 in the light of pragmatic knowledge, seemingly presumed to be *the elephant in the room* by test-developers in written language placement tests and its components, such as sociolinguistics, sociocultural, psychological, contextual, and rhetorical meanings to disclose whether they can represent functional language use in high context communications. Findings reveal that a huge gap subsists between real-life situations and given tasks in the test. At the end of the fine-grained inspection, implications and future research suggestions are stated to ensure validity and involve speaking skill more directly in YDS by emphasizing what to touch and expressly centre on the speech rather than solely assessing examinees' knowledge for recognition and appealing to multiple-choice items with the analysis of international, large-scale language tests, such as TOEFL and IELTS.

**Anahtar Sözcükler:**  
Edimsel  
değerlendirme,  
uluslararası dil  
sınavları, yabancı dil  
sınavı, yazılı söylem  
tamamlama soruları,  
YDS

**Yabancı Dil Sınavındaki (YDS) Yazılı Söylem Tamamlama Araçlarının Eleştirel Analizi: Uluslararası Testlerin Gelişime Katkıları**

**Özet:** Askeri personel, devlet memurları, lisans ve lisansüstü öğrenciler ile akademik personelin çoğunlukla katılım gösterdiği Yabancı Dil Sınavı (YDS), yerel bağlamda adayların İngiliz dili hakimiyetini belgelemek amacıyla Türkiye'de yıllardır uygulanan en popüler ve en prestijli standardize edilmiş başarı testidir. Bu çalışma, 2015-2019 yılları arasında YDS'de adaylara yöneltilen çoktan seçmeli, yazılı söylem tamamlama sorularının genellikle yazılı dil seviye belirleme sınavlarında göz ardı edilen edimsel bilgi ve bileşenleri olan sosyal dilbilim, sosyokültürel, psikolojik, bağlamsal ve söz bilimsel anlamlar bakımından yüksek bağlamlı, etkileşimli durumların ne kadar işlevsel dil kullanımını yansıttığını açığa kavuşturmak amacıyla tasarlanmıştır. Bulgular gerçek dil kullanımı ile testte verilen sorular arasında belirgin bir farkı işaret etmektedir. Çalışmanın sonunda, yapı geçerliliğini sağlamak, sınavda gerçeğe daha yakın konuşmaları dahil ederken adayın sadece bilgiyi tanınmasını ölçmek ve çoktan seçmeli sorulardan sonuca varmasını sağlamak yerine hangi noktalara değinilmesi gerektiğini uluslararası geçerliği olan sınavların detaylı incelemesi sonucunda listeleyen önerilere, gelecek araştırmalara yön vermesi için yer verilmiştir.

## 1. Introduction

Foreign Language Examination (YDS) in Turkey has strategic prominence since test-takers have been recruited, promoted, or they obtain academic title, and get a rise in salary after certifying their English language proficiency. Even so, such a decisive test has been operationalized, focusing only on grammar and vocabulary alongside reading while excluding speaking skill, although oral proficiency assessment has been conducted even at school of foreign languages in universities via exit exams and English language proficiency tests carried out at the beginning and end of the academic year. To compensate for this deficiency, five discourse completion tasks are included in YDS. However, very little critical analysis has hitherto been done to examine whether they reflect real-life communication. The mainstream research on YDS in Turkish context chiefly focused on washback effects (Akpınar & Çakıldere, 2016; Kutlu, Demirogları & Demirogları, 2020; Külekçi, 2016; Özmen, 2011; Polat, 2019, 2020), contributions of technology and internet to its preparation (Saritepeci, Duran, & Ermiş, 2019), test-takers' learning attitude, motivation and anxiety (Zeybek, 2015), academicians' willingness to communicate in English (Çetin & Kılıçkaya, 2020), its overall evaluation (Akin, 2016), analyses for state employees (Gür, 2012; Demir & Genç, 2016) and of reading comprehension paragraphs (Kıray, 2015) in addition to the validity of translation items (Dinçer, 2019). As is seen, to date, no studies on YDS have vetted the pragmatic aspects of language ability within certain linguistic contexts. To fill that niche in the literature, written discourse completion tasks (WDCTs) involved in YDS exams as the best representative of speaking competency in the test were examined in terms of pragmatic knowledge. For this reason, after a concise summary of YDS in terms of its content besides the pragmatic knowledge in the language and tests, the article aims to discuss whether WDCTs in YDS overlap with the real-life communicative tasks and to what extent language knowledge components are covered in these tasks.

### 1.1. YDS in Turkey

YDS is prepared approximately in twenty foreign languages with five options per item of 80 multiple-choice (MC) questions, and 1.25 point is given to each correct choice without deducting any points for the wrong answers. YDS was administered bi-annually (in spring and fall) until 2018; still it has been held tri-annually by Assessment, Selection, and Placement Centre (OSYM) from then on. Test construction is composed of separate sections, such as cloze test, translation, sentence and paragraph completion, fill in the blank, paraphrasing, reading comprehension, WDCTs, and odd-one-out.

OSYM (2016) puts forth that test and item-developers prefer British English, authentic data from real-life phenomena, and bear in mind the far-reaching effects of global lingua franca on non-native speakers. However, even though the validity and the equivalence of YDS and international language exams have been approved by OSYM provincially, this system does not provide an equal footing for all test-takers in real terms. This is chiefly thanks to the fact that four basic language skills are soundly assessed in internationally accepted large-scale exams, such as the Test of English as a Foreign Language (TOEFL), the International English Language Testing System (IELTS), whereas the measured values in YDS are quite restricted. In this respect, in order to mirror the basic differences between YDS and the high-stakes language proficiency tests in a systematic way and to explore how these international exams can contribute to the enhancements of YDS, a deeper descriptive and critical analysis with well-grounded claims is required in the following phases of the study.

## 1.2. Pragmatics and Pragmatic Knowledge in Language

Recently, the phrase 'less is more' seems to have been adopted by quite a few people as an inspiring life motto either to facilitate communication or give interlocutor pause, and this is to impress him/her through eloquence. Notwithstanding, that may not be so intelligible for some who randomly select and utter words or transform the message tactlessly. As a result, it might cause communication breakdowns, such as making hearer take comments amiss or rhetorical vicissitudes. Then, speakers need to establish conversational frames, etiquettes, and rituals considering cultural and social norms, conversational skills, and locutions so as not to subvert the rapport. It implies the fact that conversational flow can only be maintained respecting the embodiment of linguistic forms in social interaction besides the negotiators to build intelligible context-aware speech.

Pragmatics, as a subfield of linguistics, is concerned with attributing meaning to communication acts depending on interactional rules, manners of speakers, the social context in addition to literal aspects of the language. Pragmatic knowledge is hence indispensable for language users to conform to linguistic and non-linguistic patterns and decode context-bound codes and meanings. As such, its intrinsic character that incorporates situational and linguistic contexts in construing meaning precisely features some cardinal notions to attain language ability. One of those substantial items is speech acts, which reveals that language competency is more than excelling vocabulary and grammatical knowledge, but it also refers to the functional organization of an argument. Austin (1962) discussed this term regarding felicity conditions and then classified it into three categories as locutionary, perlocutionary, and illocutionary act, which reverberated in pragmatic studies by virtue of the underlying meaning of statements and force. In the forthcoming years, though Searle (1969) extended illocutionary acts into five groups and identified direct and indirect speech acts to emphasize the functional capacity of language, he must have overlooked a set of attributes in social interaction given that missing points appeared in orienting language users to relevant contexts to extrapolate meaning (Thomas, 1995). After Grice's maxims of conversation and theory of context-dependent and context-free implicatures to unveil the hidden meanings of propositions, politeness theory by Leech (1983) emerged owing to speakers' negligence of cooperative principles to prevent face-threatening circumstances while putting their points across.

Another issue worth addressing is the divergence between pragmatic and semantic meaning. Yule (1996) clarified their nuance by describing the former as having a strong connection with the language use and user without counting social variables to a large extent, whereas the latter encompasses linguistic units and their notable functions in communication. As his post-structuralist lens became a matter of profound debate, new prospects have continued unabated throughout the past decade. As an example, Jaszczolt (2002) and Purpura (2016, 2017) upheld the notion that pragmatic meaning was superior to the other due to its requirements of additional information in particular contexts. All the same, the scholars finally reached the consensus that was strictly compartmentalizing those concepts would be beyond the realm of possibility.

## 1.3. Changing Views toward Communicative Language Ability

Language ability, which is the backbone of pragmatics, is a multifaceted concept in terms of involving both linguistic patterns to convey communication and assigning meanings to speech according to its situational use. In other words, it allows learners to refine their

language knowledge through metacognitive strategies while making conversational inferences (Bachman & Palmer, 1996; Purpura, 2004). Considering its critical role in pragmatics, tracing back to Communicative Language Ability (CLA) in linguistic paradigm would avail to expose changing standpoints in history.

Initially, language ability was illustrated by Chomsky (1965) in respect to performance and competence. To put it differently, it was dichotomized as linguistic performance and linguistic competence to stress the difference between linguistic conformity and language use in communication aside from resolving contextualization issues and casting light upon proficiency in depth. To fill this deficiency, Hymes (1967) coined the term communicative competence and underlined the convenient use of context in cooperation with grammatical knowledge. It means that he underscored contextual factors leading interlocutors' linguistic preferences in real-life circumstances, which was reinforced by Halliday and Hasan (1976, 1989) in that context would attach importance to decipher meaning in conversations; hence language cannot be regarded as fundamentally structure-based. Afterward, Halliday (1994) introduced a systematic functional perspective through pragmatic and semantic meanings, which remain more significant when compared to language forms regardless of the fact that they are intertwined and complementary to one another in linguistics. In the same vein, Canale and Swain (1980) propounded a model with four dimensions underlying the efficacy of context. Grammatical competence (decontextualized knowledge) as the linguistic property of dialogues, discourse competence indicating syntactic and lexical proficiency of language users in specific contexts, sociolinguistic competence signaling basic knowledge about productive communication with speakers, discussion topic, context, and social factors, and finally, strategic competence in the role of finding a way out to communication problems were all covered in their standard. However, Canale (1983) reformulated this design, and over-riding attention was attached to language use in a variety of sociolinguistic contexts. Despite his attempts to put the speaker's understanding forward in interpreting meaning within a relevant context, no affiliations among the four afore-cited constituents can be discovered. This must stem from the fact that they separated pragmatic items inapplicably rather than allowing them to overlap systematically.

Taking the setbacks of former examples into account, CLA was mostly associated with pragmatic elements of language, then strategic competence and language knowledge were outlined by Bachman and Palmer (1996) as integral parts of the model. Strategic competence was principally related to metacognitive strategies, whereas language knowledge comprised of pragmatic ability and organizational knowledge, such as rhetoric, grammatical ability, and functional aspects of language use, respectively (Tadayon & Ravand, 2016). Similarly, Fulcher (2003) delineated language ability in five categories, including language competence, strategic capacity, textual, pragmatic, and sociolinguistic knowledge. Nonetheless, as criticized by Purpura (2004), these researchers underrated hidden layers of meanings in propositions besides neglecting linguistic patterns to be credited while expounding pragmatic meaning in spite of its contextual-based language lens; it follows that they were cognitive-oriented instead of communication-based systems as in *the trait* point of view. Accordingly, highlighting the weight of meaning and holding other designs (Bachman & Palmer, 1996; Canale & Swain, 1980; Celce-Murcia, Dörnyei, & Thurrell, 1995; Fulcher, 2003; Halliday & Hasan, 1976, 1989; Hymes, 1967), Purpura (2004) extended pragmatic ability in language use offering a variety of sentential meanings in high context situations, such as literal, intended and implied meaning of utterances. Furthermore, the pragmatic component in his revised model was fairly extensive on account of other essential elements it encapsulated, such as contextual, sociocultural, sociolinguistic, psychological, and rhetorical meanings. This model

has stayed one step ahead of the others in that speakers' grammatical proficiency, construing the conversation within a restricted frame, and eliciting their pragmatic competence to make inferences about the context were appraised contrary to speech acts.

#### 1.4. Brief Analysis of Pragmatic Knowledge Components

Having recognized that pragmatic aspects of language knowledge have been of great significance as the linguistic units in language ability models, some researchers rearticulated the salience of pragmatic features in the models to reflect the ability of learners in real-life cases. Yet, before scrutinizing pragmatic knowledge in tests and its operationalization in studies, a concise analysis of its components would be crucial.

In a broad sense, sociolinguistic meaning refers to the awareness of how to make contacts through language in a society respecting a good number of factors, such as social meaning, register, norms, cultural and social identity markers, genre, and so on. At this point, social meaning needs to be specified with some sociolinguistic variables determined by Brown and Levinson (1987). They enhanced the politeness hypothesis (Leech, 1983) and defined a tripod with power, distance, and (rank of) imposition. Thus, the interlocutor's communication strategy would be settled to balance politeness and linguistic patterns in order to apprehend implications in that context. The speaker's eschewing face-threatening cases or utilizing politeness strategies felicitously would be the signs of his/her pragmatic knowledge. Otherwise, as Thomas (1983) described, communicative conventions in L2 and pragmatic force, such as performative, directive expressions might be misunderstood, or intended meaning cannot be transferred to the interlocutor accurately owing to their incorrect applications, direct translations, or falling short in catering to contextual cues (Gumperz, 1982). Then, speech acts between a native speaker and a second language user might turn into a discordant conversation creating pragmalinguistic failure(s), which will reveal pragmatic incompetency of the L2 speaker (Taguchi, 2018).

As for sociocultural meaning, it has to do with the familiarity of learner's social environment, and its cultural impacts on his/her behaviour, psychology, speech act, the manner in speaking, and maintaining a successful interaction in a speech community (Lado, 1961). As such, mastery in verbal expressions, circumlocutions, semantic opaqueness through literate, illiterate meaning, modality, and language forms besides applying to cultural expectancies would help keep natural language understanding. Hereof, the model developed by Hymes (1974), who signified contextual characteristics as situation, participants, ends, act sequence, key, instrumentalities, norms, and genres (SPEAKING), is to be cited since it correlates social, psychological, and physical elements of context from a sociocultural perspective. Otherwise, the sociocultural barriers in a specific discourse may lead to sociopragmatic failure(s) in cross-cultural communications due to the inadequacy of cultural customs or sociocultural knowledge in the target language and misinterpretations of speech actions.

As to the other agents, the psychological meaning is concerned with the speaker's position, *mood, attitude, feeling, and disposition* (Ochs, 1996, p. 410), the usage of grammatical forms in identifying the tone and gesture, intonation throughout an on-going speech. Contextual meaning dwells on language use in interpersonal framework dealing with their face, humour, identity, gender, and politeness (Locher & Graham, 2010), though. Finally, rhetorical meaning points out the genre, organizational mode, and coherence. In the light of the above-mentioned maxims of pragmatic knowledge, it can be straightforwardly asserted that pragmatic appropriateness in L2 demands a number of issues to be concurrently treated for

each proposition. (Cohen, 2019, Dewaele & Dewaele, 2020, Taguchi, 2018; Timpe-Laughlin, Green, & Oh, 2021).

### **1.5. Pragmatics in Tests**

In the past, the assessment of second language ability used to focus on structural form, emphasizing discrete-point approach via one single dimension of abstract language use without adopting any particular context. Subsequently, Lado (1961) laid the foundations of his Contrastive Analysis (CA) hypothesis in linguistics on the basis of behaviourism disdaining the creativity issue in the language (cf. Chomsky, 1957). As a result, Carroll (1968) counted in communicative context to measure language competency of L2 learners upon reorganizing the former model. Thereafter, Spolsky (1968) reiterated the weight of creativity and affirmed that an all-encompassing system with tenets subsists in language competency. Notwithstanding, Oller (1979) expostulated dividing principles of language proficiency into distinct parts, which would be incompatible with the unitary competence. He set forth both internalized grammar and pragmatic expectancy grammar to contribute to language processing and interpreting meaning. Thus, he initiated the conceptualization that pragmatic components and linguistic forms ought to be harmonized to impede the hegemony of skill-based assessment, unilateral language use through linguistic structures. He instigated a nascent liaison between integrated skills and functional aspects of language use in tests.

The tendency of incorporating integrated practices in language tests prompted the meaning construction of test-takers, and increased authenticity, which would not be probable with indirect assessment tools, or artificial tasks (Clark, 1975; Jones, 1985; Roever, 2006). In other words, real and context-specific tasks encouraging interactive communication were to be offered to test-takers. For instance, Hudson, Detmer, and Brown (1992, 1995) aimed to enlighten pragmatic failures of examinees through sociolinguistic variables in cross-cultural contexts besides tackling variability and reliability problems in tests. Yet, some unforeseen answers limiting generalizability sounded psychometrically unreliable results due to MC WDCTs in particular, which was chiming in with Brown (2001), Cohen (2019), Liu (2006), Roever (2018), Yamashita (1996), and Youn and Bi (2019). Their research accentuated the fact that MC items were erroneous and tricky owing to the translations between the target language and mother tongue of test-takers, and measurement of pragmatic competence might be problematic thanks to contextually-bounded nature and the difficulty of checking sociocultural and sociolinguistic meanings via these question forms. Additionally, their results documented that analysing pragmatic elements apart by classifying them into unique subgroups would potentially weaken functionality, discredit validity in test construction, and marginalize the operationalization of pragmatic knowledge in overall second language ability.

### **1.6. Pragmatic Language Knowledge in Large-Scale English Language Proficiency Tests at International Level**

Some large-scale high-stakes standardized tests, such as TOEFL, IELTS, Test of Spoken English (TSE) superseded by TOEFL, take place all over the world to specify the language level of test-takers. The speaking section in TOEFL and IELTS embraces the context-dependent cases, which prompt test-takers to render meaning through reciprocal conversation. These realistic practices include dialogues, independent speaking tasks, integrated tasks with other language skills inside or outside the class, monologues (i.e., individual long turn), and discussion with test-examiner. As such, they portray test-takers' decision-making process as well as their simulation-driven product development, which is

almost unfeasible in traditional paper and pencil test format (Bachman, 2002). Furthermore, the tests account for the reactivity of test-takers since response expectancies would be identified with immediate feedback during the speech, and the way of providing input, ensuing answers along with the setting (Bachman & Palmer, 1996). The interactive tasks concomitantly gather psychological, situational, and social aspects of language within contextualized communications. Dissimilar to traditional assessment, those authentic tasks in speaking tests reveal pragmatic language knowledge in that they allow test-takers to use informational structures along with extra-linguistic features, contextual and paralinguistic cues to manipulate meanings in cognitive and heuristic sense and convey the message to the hearer. Then, sequential negotiation between interlocutors will be measured in the light of pragmalinguistic, sociopragmatic entities.

Assessment of pragmatic knowledge in communicative tests would also prove the validity of the given expressions. Inclusive and collaborative practices with grammar and pragmatic aspects of language ability will illustrate to what extent the test is valid in assessing the speaking skill of test-takers besides reflecting the severity of understanding language ability in general. Another indication of validity in these language tests must be the variety in tasks and speech acts in cultural and cross-cultural contexts, such as request, thanking, apologizing. In respect to reliability, as affirmed by Hudson et al. (1992), tests also search for the measurement of errors more than one-way error analyses refraining from overgeneralization technique and labeling all of them as ‘random.’ Additionally, estimating degrees of formality, power, politeness, speech continuity, accuracy and fluency, and pronunciation to discover L1 influence, and examining the use of discourse markers and cohesive devices to check the logical sequence between sentences would only be probable via employing real or real-like tasks. As a result, foregoing proficiency tests must be thought of as models while designing all sorts of high-stakes language tests.

## **2. Purpose and Methodology of the Study**

This study was set out to explore whether WDCTs in YDS-English reflect real-life communicative tasks through the lens of pragmatic knowledge and to what extent language knowledge components are involved in the test. Accordingly, two research questions were developed as follows:

1. Do WDCTs in the YDS exam overlap with the real-life communicative tasks from the pragmatic aspects?
2. To what extent are the components of language knowledge covered in WDCTs in the YDS exam?

### **2.1. Research Design and Data Collection**

Initially, WDCTs in the foreign language examination were defined as the scope of the study since they would best represent the speaking competency and indicate test-takers’ potential oral proficiency through the MC questions out of all question types. Then, all of the WDCTs within the last five years’ YDS questions were aimed to be incorporated into the scrutiny; nonetheless, only 45 WDCTs from 2015 to 2019 exams could be investigated from pragmatic dimensions due to the recent policy of OSYM that only ten percent of the test could be accessible to the public. Firstly, 9 YDS-English exams were found on the official website of OSYM. Thereafter, they were downloaded, then WDCTs were extracted from the other sections and categorized year by year. This analysis primarily regarded assessing the functionality of WDCTs in terms of covering real-life communicative tasks in light of

pragmatics and language knowledge. In other words, after searching for the exam questions on the net, the steps of appraisal and analysis would be subsequently conducted.

In order to give a clear portrait of pragmatic components of WDCTs systematically, a checklist was developed and tabulated (see Table 1) only after a comprehensive literature review and the revision of CLA models (Bachman & Palmer, 1996; Purpura, 2004, 2014, 2016, 2017; Suesattayawong, 2006). The researcher created the checklist without conforming to one specific model or list, taking into account the fact that each principle was detected to overlap to certain extents despite being characterized as single segments. Thus, before the analysis, the expert view of a professor in the English language teaching department was obtained to provide evidence of face validity, feasibility, and the utility of this instrument with a comprehensive list of the pragmatic components. In the end, the identified pragmatic dimensions to evaluate the strengths and weaknesses of the tasks were ‘appropriate language use,’ ‘cultural meaning and reference,’ ‘social and cultural norms,’ ‘strategic competence,’ ‘direct expressions and negotiations,’ ‘representativeness of task,’ ‘emotional and attitudinal stance,’ ‘extra-linguistic knowledge,’ ‘textual information,’ ‘power, distance and imposition,’ and ‘social identity markers.’

## 2.2. Data Analysis

Based on this checklist, the pragmatic features were identified, and their occurrences were ticked. In this way, the researcher also realized to what extent language knowledge was thoroughly treated in these tasks. To find detailed answers to the research questions, the researcher principally used discourse analysis in the meaning-making of the interlocutors’ speech in the dialogues, particularly focusing on the linguistic structure, sociolinguistic meanings, psychological meanings, grammatical meaning, and sociocultural meanings. As suggested by Clennell (1999), Hutchby and Wooffitt (1998), and McCarthy (1991), while investigating the design of expressions, their location with the turn-taking mechanism in the interaction, and unveiling the language ‘behind the sentence,’ this broad field of study is of a pivotal role.

Moreover, after the researcher completed her first examination, two colleagues, one non-native and one native (American) instructor, assessed WDCTs adopting the checklist to ensure inter-rater reliability of the study. Then, she and the two raters cross-checked their provisional findings and discussed some WDCTs together once more to eliminate the confirmation bias. Finally, they ironed out the kinks, and this also helped the researcher interpret the tasks much better in terms of the included components of language knowledge.

## 3. Findings

Table 1 presents the concise analysis of WDCTs in YDS from the pragmatic aspect and it indicates that in the first phase of 2015 YDS, stating a background seems to be regarded in three of five questions, which makes test-takers to be involved in the dialogue inherently. Furthermore, the inclusion of explicit cues to display persistent manner (e.g., *sometimes, always*), quantifiers (e.g., *much, more*), multi-word adverb phrases, and modifier words as adjectival phrases (e.g., *really convinced*) besides contextualization cues of lexis (e.g. ..., *but...*) were painstakingly made. Interlocutors appeared to, *inter alia*, acknowledge their positions, avoided threatening faces by keeping the mutual agreement to maintain a friendship frame with the exception of only one dialogue. In addition, coherency and lexical cohesion of the conversations were accomplished throughout the text. Nevertheless, speakers mostly asked and answered the questions explicitly, ignoring implied, intended meaning types and indirect



expressions, which was a requirement, particularly in the second speaker's utterances due to not declaring a conventional example to elicit the issue. Thus, it would result in losing the nature of interactional language forms in the general sense. As to the topics, specific terms from the literary world, such as genre, pragmatic, fiction, and stereotypes (e.g., *technology, media*) were incorporated to solely check test-takers' vocabulary knowledge. Moreover, reading comprehension questions seem to be prepared even without presenting a valid picture of realistic linguistic entities, such as using *wasteful* rather than *extravagant*, *s/he* instead of clarifying the person with *he or she* in the text, *to* than semi-fixed phrase *in order to*, *but to yet* or *where you are* contrary to *where you are positioned*.

Table 1.

*Rhetorical Structure of WDCTs in Foreign Language Examination*

Pragmatic Dimensions of Questions and Answers	Exams								
	2015 (1)	2015 (2)	2016 (1)	2016 (2)	2017 (1)	2017 (2)	2018 (1)	2018 (2)	2018 (3)
Background	√		√					√	√
Appropriate Language Use	√		√	√			√		√
Cultural Meaning & Reference									
Social & Cultural Norms									
Strategic Competence	√	√	√	√	√	√	√	√	√
Direct Expressions & Negotiations	√	√	√	√	√	√	√	√	√
Representativeness of Task			√						√
Emotional & Attitudinal Stance									
Extra-linguistic Knowledge									
Contextual Cues	√	√	√	√	√	√	√	√	√
Textual Information	√	√	√		√	√	√		√
Power, Distance & Imposition			√			√	√		√
Social Identity Markers			√			√	√		√

In respect to the other test in autumn in 2015, presupposition with the words *exactly* and *definitely*, hesitation markers (e.g., *oh, well...*), idiosyncratic treatment of quantifiers, modifier adjectives (e.g., *beautiful city, good example, certain aromas*), connotation (e.g., *make a decision*), phrasal verbs (e.g., *look for, dream of, back up*), indirect signals *if* and *may-might* to allude to the stretchability along with interpersonal functions by building or keeping relationships in L2 constituted the symbols of pragmatic competence to a degree. Moreover, sentence markers (e.g., *I don't think, I think*) from real-world corpora, expressions strengthening the decisive stance of the interlocutor, and increasing fluidity (e.g. ..., *right?*) were legitimately tapped to generate hypotheses about their intentions and schematic expectations. Still, the analysis of American native speaker revealed that the dialogues did not sound natural enough owing to inconvenient word choices and syntactical patterns, such as ...*requires more thought, for instance* (a mismatch with informal talk), *I don't think it makes...* (*it will make* would be more acceptable

for Turkish speakers). Additionally, overstatement of the pronoun 'I' (e.g. *...until I was .., I was.., but I .., I ...*), fairly obvious correct answers without strong distractors (e.g., *I (don't) think* and *as I*) requiring no pragmatic competence, and embellished propositions as hasty suggestions (e.g., *Please, let's..*) were asked to test-takers in spite of the necessity of verbatim representations of statements, and non-verbal communication, such as gestures, facial expression or body language. These would all enunciate the artificiality of those so-called authentic materials.

The first exam in 2016 released a background and reflected daily-language through informal words as phrasal (e.g., *take away, cool off*), attributive verb (e.g., *agree*-instead of *say*-), figurative form (e.g., *leave a mark*), formulaic expression (e.g., *what do you mean?*), common vocabulary use (e.g., *stir*), downtoner, collocation (e.g., *as far as I know*), hedges, boosters (e.g., *but, sure*) in the role of pre-announcement of the forthcoming face-threatening case, comparative quantifiers (e.g., *faster, higher, cooler, smaller, harder, more*), the epistemic and evidential form of elicitation (e.g., *must*), incipits as the minimalist, succinct response (e.g., *Ok. But,..*), interactional form (e.g., *probably*), and mid-utterance particle (e.g., *you know*) to preserve the solidarity and resolutions (e.g., *ok, so*). Besides, the last dialogue between student and professor symbolized social distance, power, imposition, or social deixis with the determination of politeness and strengthened psychological stance (e.g. *..., right?*). On the other hand, though widely featured collocations, attributive verbs, and phrasal came to the fore in the text as sociolinguistically convenient, their probability of being correct answers was nearly unfeasible in terms of the pragmatic quality.

The second YDS in 2016 appears to have pragmatic ambiguity since the doctor's utterances to the radio presenter were neither relevant nor intelligible to address the claim (e.g., *analyzing brain in labs, its donation, people participating in a project*). No greeting patterns or opening lines were included to purport psychological meanings or feelings either. Besides, the awkward form of expressions cannot be left unattended herein due to the selection of words for effective communication, such as *persuade to agree to donate* than *persuade to donate*, which makes the audience presuppose strained interaction throughout the conversation. Still, the doctor in a dialogue opted for using 'we' to build solidarity sociolinguistically and endeavour to be in alignment with other colleagues psychologically. In the sequence that follows, grasping the meaning of 'what does it take?' would explicitly serve as a lead-in for test-takers without the construction of any judgments. Yet, this reply would not be favoured in this context thanks to its hardly indigenous linguistic forms as was articulated by the native rater. In sober fact, some semantic and syntactic problems subsist in the text in spite of the utilization of English or American names along with informal idiom (e.g., *bet on*), complaint-like phrases (e.g., *it doesn't make sense*), mitigating language to compromise (e.g., *you may have a point*), frequent mental two-word cluster (e.g., *I believe*) positioning interlocutor on an epistemic scale measuring the certainty of commitment, tag question as hedging in softening function and hints of negotiation, a suggestion with *may* as an illocutionary force, and interrogative request (e.g., *can you...?*) demanding sequential arrangement as an indirect expression of speech act indicating politeness. Firstly, pragmatic ambiguity appears again with negative premises (e.g., *I don't think...means ...a lot, do you? – yes, I do*). Then, applying one-word answer (e.g., *layman*), unparalleled utterance (e.g., *Ok* right after the question 'what do you mean?') requiring tone to be decoded, and defective organization of correct answers depending only on direct or indirect solicitations to elucidate the content integrity (such as *what do you mean, what is it, why, can you explain...?*) were all detected to be the failures.

In the first round of 2017 YDS, test-takers were introduced to dialogues in which epistemic words were embedded as sentence markers (e.g., *I guess*), contextualization cues (e.g., *oh, I see, well*) to downgrade the certainty of face-threatening in communication and appeal to other interlocutor's positive face in sociolinguistic context. Moreover, the conditionals, direct (e.g., *I don't agree...*) and indirect utterances (e.g., *can you..?*) with presuppositions, were integrated into the test. On the other hand, test-takers would necessitate only lexical knowledge and word awareness (e.g., *indoor, attic, basement, balcony, sunlight*) besides noticing some specific linguistic concepts, such as negative connotation or synonym to identify the answer. That is, the test will not lead examinees to any indecisiveness or hesitation in selecting the correct answer, nor will it demand them to be knowledgeable about the essential points in pragmatics. As a result, more indirect phrases, such as *I suppose, I reckon, I assume* were in need to shy away from opinionated manner, imposing restrictive notion and violating opinion-reticence maxim (Leech, 1983). Moreover, meticulously adapted communicative devices enacting test-takers to identify indirectness in terms of psychological divergence in the social context, conversational implicatures, and entailments were to be conveyed to the target audience.

All of the interlocutors in the second phase of 2017 YDS were European (Ann, Jane, Robert, Carl). The dialogues were between the patient-psychologist, journalist-anthropologist, journalist-surgeon to make us presuppose politeness in an external context. Especially in the first speech, backchannels, prosodic cues, such as intonation contours of some phrases (e.g., *extensive, while on a horse, only, first real evidence*) were expected in tasks to reveal attitudinal stance due to lack of greetings or preludes in the beginning and to provide smooth transitions among the propositions. The social norms, addressee's anticipated agreement in an assumed sociocultural context would successfully contribute to determining '*traditional perceptions*' as the correct answer in addition to its rhetorical appropriateness to the context. Nonetheless, the following tasks seemed to be based solely on a one-word answer, such as *interval*, the continuity between wh... question and metaphoric definition of the verb (e.g., *get*), the semantic affinity of '*come to a decision*' as a collocation, '*beneficial*' as a modifier, and finally the consecutive notions in the form of paraphrases supporting the obscurity in the text. It seems as if they all aimed to absolve test-takers from the legitimacy of reasoning. To put it the other way around, WDCTs, as the most salient part of the test representing real-life communication, unveiled ellipted interactional forms in different social situations entailing an incomplete sociolinguistic profile in high context.

As for 2018, nearly all of the talks were formal and conducted by an English reporter, a researcher, a journalist, a professor, and friends to signify distance. Some obtrusive points impeding pragmatic fluency ought to be treated pre-emptively so as to nail down to what extent the examination allows test-takers to extrapolate those quasi-factual narrations. The dietician-reporter dialogue was discovered to be mechanic in that almost no signs of authenticity can be found throughout the speech considering varying degrees of direct statements, imperative sentence (e.g., *tell us..*), the dearth of mitigating phrase, hearer-oriented requests, and most importantly pragmatic ambivalence in reporter's second delivery for anchoring negotiation since no benefits of using this *blend* and *health* were mentioned. In what follows, the sudden and unexpected involvement of the reporter and his/her behaviour's being the target point led us to examine the questions which seemed to measure only articulately expressed items according to the comments of documentarist rather than searching for a layering effect of meanings. Similarly, the dialogue between journalist and geophysicist would expose primary focus on talk with explicit enunciation regarding interpersonal meaning in contextual terms; otherwise, *key questions* in A could also be taken

as gospel. Another elusive and dissatisfying answer, which would be improbable to predict unless it was given in the options, was introduced with the conversation between a professor and a student. Still, word smart test-takers must have traced *vulnerable*, *predator*, *unsafe* to find the correct answer, or a little mastery in reading comprehension would be adequate. In conclusion, the negligence of the exam board in considering pragmatic knowledge and discourse analysis in the exam can be easily noticed.

As in the former one, test-developers intentionally gave European names to speakers in the second YDS in 2018 and maintained a delicate balance between role identities with only one journalist-psychologist dialogue. Notwithstanding, serious problems were recorded to manifest themselves after a painstaking analysis. As was aforementioned, discontinuity in discourse processing and misguiding the addressee were assumed to reiterate on account of the fact that vague language was encountered notably in the last conversation. Carole's two explications cannot be asserted to overlap with one another in that the first utterance did not refer to any other person to share goals with, whereas the need of an interrogator was recognized there and then in her second delivery. In the same vein, Sarah's explanation could be sustained with the third option of all the alternatives, but for its overgeneralized claim allowing the flow of logical argumentation and face considerations. Yet, test-takers might lose focus in view of their weaknesses in the acumen, cultural differences (e.g., *university* instead of *college*) apposition markers frequently utilized in informal speech for the claim augmentation (e.g., *that is to say*, *in short*), lack of precursors- *pace changes*, *uptalk*, *stress or elongation*-, promptness etiquette, and the difficulty in deciphering the psychological meaning of *backfire* or associating the personality with foregoing adjectives.

Although identical structures and functions were implemented in the last 2018 YDS, some salient points were to be emphasized considering pragmatics. First and foremost, narrations were generally supported with a background in particular contexts, such as exam revision, reading a magazine article and learning the number of obese people, and university enrolment. Dissimilar to some other questions investigated above, the second, fourth, and fifth ones were factual and more pertinent to be accepted as native-like performance in terms of making pragmatic competence functional. In reference to the other two conversations, straightforward logic (e.g., *giving birth-mother*, *individual-collectively*) and mediocre reading skill would be enough to expound pragmatic dimensions in the text.

After a rigorous examination of WDCTs, the prevalence of linguistic entities, strategic language use in terms of activating metacognitive strategies, and contextual cues would stand out. Nonetheless, they can be just put on the far side of pragmatics, and it is reasonable to denote that these tasks involve a blind bit of communicative competence in the grand scheme. Besides, sociolinguistic appropriateness in the analysis merged under the name of contextual cues, partially social norms, and social meanings, such as politeness, yet they did not embody the social identity markers extensively. The most problematic sides were correlated with sociocultural, contextual, and psychological meanings. These points underrepresented pragmatic knowledge in YDS on the whole, in addition to the non-isomorphism between linguistic structures and semantic structures, which culminated in adopting ill-equipped tasks to perform professional, and real-life acts.

#### 4. Discussion

Having identified pragmatic problems in 45 questions, and considered international exams with a pioneering role in eliciting pragmatic language knowledge, WDCTs were explored to

be largely mechanic and implemented with strained interaction through stereotypes without originality or variety in speech acts, which is expressly required in social norms. This result is in tune with Brown (2001), Hudson et al. (1992, 1995), and Yamashita (1996). Task-oriented scenarios hardly resemble real interactions besides the dearth of stronger distractors and more legitimate alternatives to the given situations in MC items (Beebe & Cummings, 1996; Golato, 2003). Additionally, language assessment of test-takers was largely administered using strategic competence and metacognitive strategies despite being notoriously devoid of sociocultural appropriateness as was in Canale and Swain (1980) who also opted for narrowing the testing context than espousing the multicomponent view. In other words, test-developers appeared to design YDS without respecting the interplay of cognitive knowledge with the contextual competence of test-takers, and hence they did not bridge the gap between modern language ability models and the ways of measuring L2 pragmatic knowledge according to theoretical underpinnings (Chapelle, 1988). However, test items and tasks based on contemporary and theoretical language modeling were to activate cognitive, communicative, and social skills conforming to the interactionist approach so that test-takers would thoroughly assimilate and create pragmatic meanings as in international language tests. Correspondingly, dissimilar to IELTS-takers or TOEFL-takers, the examinees must have experienced difficulty in skill areas due to being linguistically unprepared (Kamaşak, Sahan & Rose, 2021).

In addition, although the answers were mostly established in a coherent framework on the basis of what and why questions and demanding further explanations about the topic, any mitigated disagreements or schismogenesis (Bateson, 1935) were not encapsulated in spite of their usage in real-life. Policymakers and test-developers cannot have counted that speech acts usually occur only after some turns (Korsko, 2004) than one or two consequent replies, and similarly, one language function to be interpreted within measurement design cannot straightforwardly draw on assessing the comprehension or production of pragmatic knowledge. Tasks were not cultivated with drama (Grabowski, 2009), paralinguistic cues, intonations or phonological forms (Purpura, 2004; Yamashita, 2001), emotions, such as laughter, backchannelling, implicatures, gestures (Beebe & Waring, 2002; McNamara & Roever, 2006). As a result, test-takers cannot even monitor false starts or dysfluency in speech.

On the grounds of abstract expressions based on linguistic structure or grammatical meaning, simple turn-taking mechanisms, and interference interactions besides the threats to validity of non-interactive tasks, WDCTs-MC items cannot be reported as successful agents in measuring pragmatic competency (Kasper & Rose, 2002; Roever, 2006; Yamashita, 2001). To illustrate, scripted replies in controlled tasks, which were unrelated to productive skills and pragmatic language use, cannot depict speakers' and hearers' input to be interpreted later (McNamara & Roever, 2006), nor can they offer an opportunity to present justification of the behaviours or responses.

To put it simply, lexical cohesion and rhetorical practices in addition to fixed social norms, negotiators, contextual cues, unilateral thought processes with meta/cognitive strategies, and sociolinguistic variables were explored in WDCTs in a general manner. Even so, social distance, determination of power, taking the risk of imposition of utterances along with prototypical language use, juxtapositions, resolutions of conflicts, and different face-threatening cases were not included properly due to superficial semantic processing or absence of native speakers or test-partners in the exam. Thus, discourse deixis and tools for strong interaction to record the assessment of sociocultural ability cannot be detected as in

the analysis of Korskó (2004), which must have blocked their involvement and assimilation of the tasks. Finally, WDCTs in YDS include neither extraordinary phrases displaying emotional and attitudinal stances of interlocutors nor interactive contexts in line with the purposive domain. Therefore, these items are conspicuously insufficient to fulfill communicative functions different from the ones in the international language tests.

## 5. Conclusion and Implications

The study indicated that dividing the pragmatic competency of test-takers from other language knowledge or comprehensively assessing their pragmatic knowledge as a salient factor in language ability was intensely disputed to date (McNamara & Roever, 2006). A wealth of research in the field adopted restricted scopes, and distinct pragmatic items, such as speech acts (Hudson et al., 1992), eliciting (see Purpura, 2004), intended meaning while ignoring meaning types to some extent (see Grabowski, 2009), and related models (Beebe & Waring, 2002) by disregarding some of the components of pragmatic knowledge. Thus, they were not planned taking into account an interdisciplinary perspective and extensions in language use. In this respect, the overall findings provide a firm basis for the assertion that the elusive and inextricable nature of pragmatics playing a critical role in making learners comprehend the utterances, leverage linguistic resources, and accomplish communicative competence in various situational contexts to formulate discourse and create syllogisms via logical reasoning accounts for the failures in L2 testing.

Taken further, resorting to WDCTs in YDS to assess the speaking skill of test-takers can be acceptable only in terms of its practicality and grammatical meanings. On the other hand, the greatest failure of WDCTs was that they could not exhibit real scenarios regarding the sociocultural, psychological, and contextual entities. Consequently, performances do not illustrate real language proficiency and pragmatic competency of test-takers (Brown, 2001; Farhady, 1983; Grabowski, 2009; Purpura, 2004; Roever, 2006). Instead, performance assessments must be favoured for the perception of pragmatic values, interpretations of communicative tasks, the interface of grammatical knowledge, and pragmatic components of a language in a variety of contexts. Likewise, specific task functions could be incorporated so that test-takers can co-construct meaning for productive performance (McNamara & Roever, 2006). Another alternative may be the reciprocal tasks or at least semi-direct tasks where implied meanings are embedded in high contexts to overlap all skills systematically (Grabowski, 2009), and introduce authentic materials considering psychological, cultural, linguistic, social, and contextual factors similar to large-scale international tests.

In the guise of conclusion, interactive speaking, such as task-based role-plays in a certain context, short discussions with test-partner (Brown, Hudson, Norris, & Bonk, 2002), dialogue with the examiner, and integrated speaking tasks should be comprised to meet communicative goals by adhering to rubrics created and graded by test-developers in light of pre-determined expectancies, and universal benchmarks as in international language exams. Then, test-takers' capacity of conceptualizing implied meanings in the interpretation process via psycholinguistic representations (Douglas, 2005) and contextual factors (Schegloff, 1987) will gauge their pragmatic ability much better than expecting simple inferences of examinees from a few speech acts, or a fixed situational structure (Johnson, 2001).

As for the implications for further research, this analysis highlights the necessity of a full-fledged test to determine the general language knowledge of examinees and reflect reliable results of their strategic and communicative language use. Furthermore, the pragmatic and

social dimension of language in YDS can be discovered by investigating test-takers' use of linguistic features and conveyance of meaning with computational linguistics, discourse analysis, or conversational analysis (Lazarton, 1992). Having identified general obstacles in implementing pragmatic knowledge, the immediate needs of learners from a wide range of educational backgrounds, and the sociocultural factors, YDS should be reformulated with non-conventional linguistic structures and context-dependent interactions to construct validity by keeping the overall findings of this analysis in mind. Another significant contribution to the field in the light of these points will be providing empirical evidence on the emotional and attitudinal stances of examinees after they have taken the up-dated version of YDS, as was asserted by Dewaele and Dewaele (2020), and Özer and Altay (2021).

The analysis also underscores that issues should be infused into sociocultural meanings, such as cultural norms, the authenticity of the text, preferences, metaphoric expressions as well as interpersonal communication, which turned out to be outstanding figures to be treated in the end (Ochs, 1996; Yamashita, 1996). That is because failures in sociocultural knowledge would accelerate pragmatic deficiency and lead to psychological convergence in turn. Moreover, the direct and transparent convention of the Turkish language misaligning with the English language would make item developers abstain from verbatim translations. Finally, the test difficulty in terms of pragmatic and grammatical aspects needs to be examined to ascertain the appropriateness of the exam for all examinees at different levels. Last but certainly not least, researchers are kindly invited to conduct more in-depth studies to unearth alternative ways to measure the pragmatic ability of test-takers or enhance the current ones addressing how to overcome the failures and enlighten its binary relation with other linguistic features. In this way, YDS will be more functional to determine the language levels of test-takers as a high-quality language assessment.

As to the limitations of the study, only MC WDCTs in YDS from the years 2015 to 2019 were incorporated into the scope. However, this time interval was a bit restricted, and the number of the tasks included in YDS was only 45. As for the instrument, adopting a checklist from the recent literature whose validity and reliability were statistically tested could be more convenient in consideration of the assessment and evaluation issues. By the same token, interviews with some of the test-takers would consolidate the findings and probably disclose other key issues to be addressed in further research.

### **Ethical Issues**

The author confirms that the study does not need ethics committee approval according to the research integrity rules in her country.

### **References**

- Akın, G. (2016). Evaluation of national foreign language test in Turkey. *Asian Journal of Educational Research*, 4(3), 11–21.
- Akpınar, K. D., & Çakıldere, B. (2013). Washback effects of high-stakes language tests of Turkey (KPDS and UDS) on productive and receptive skills of academic personnel. *Journal of Language and Linguistic Studies*, 9(2), 81–94.
- Austin, J. L. (1962). *How to do things with words*. Oxford, UK: Clarendon Press.
- Bachman, L. (2002). Alternative interpretations of alternative assessments: some validity issues in educational performance assessments. *Educational Measurement: Issues and Practice*, 21(3), 5-19.
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.

- Bateson, G. (1935). Culture contact and schismogenesis. *Man*, 35, 178–183. <https://doi.org/10.2307/2789408>
- Beebe, L., & Cummings, M. (1996). Natural speech act data versus written questionnaire data: How data collection method affects speech act performance. In S. M. Gass & J. Neu (Eds.), *Speech acts across cultures: Challenges to communication in a second language* (pp. 65-86). Berlin: Walter de Gruyter.
- Beebe, L., & Waring, H. Z. (2002). The pragmatics in the interlanguage pragmatics research agenda: The case of tone. Paper presented at the American Association for Applied Linguistics Conference, Salt Lake City, Utah.
- Brown, J. D. (2001). Six types of pragmatics tests in two different contexts. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 301-325). Cambridge, UK: Cambridge University Press.
- Brown, J. D., Hudson, T., Norris, J., & Bonk, W. (2002). *An investigation of second language task-based performance assessments*. Honolulu, HI: Second Language Teaching & Curriculum Center, University of Hawaii
- Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge, UK: Cambridge University Press.
- Canale, M. (1983). On some dimensions of language proficiency. In J. Oller (Ed.), *Issues in language testing* (pp. 333-42). Rowley, MA: Newbury House.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Carroll, J. B. (1968). The psychology of language testing. In A. Davies (Ed.), *Language testing symposium: A psycholinguistic approach* (pp. 46-69). London: Oxford University Press.
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6(2), 5-35.
- Clennell, C. (1999). Promoting pragmatics awareness and spoken discourse skills with EAP classes. *ELT Journal*, 53(2).
- Chapelle, C. (1998). Construct definition and validity inquiry in SLA research. In L. Bachman & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32-70). Cambridge, UK: Cambridge University Press.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton de Gruyter.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Clark, J. (1978). *Direct testing of speaking proficiency: Theory and application*. Princeton, NJ: Educational Testing Service.
- Cohen, A. D. (2019). Considerations in assessing pragmatic appropriateness in spoken language. *Language Teaching*, 1–20. <https://doi.org/10.1017/S0261444819000156>.
- Çetin, K., & Kılıçkaya, F. (2020). An inquiry into Turkish academicians' willingness to communicate in English. *Journal of Narrative and Language Studies*, 7(13), 289-308.
- Demir, D., & Genç, A. (2016). The analysis of foreign language proficiency exam for the state employees based on the CEFR. *Hacettepe University Journal of Education*, 31(1), 53–65. <https://doi.org/10.16986/HUJE.2015014090>.
- Dewaele, J. M., & Dewaele, L. (2020). Are foreign language learners' enjoyment and anxiety specific to the teacher? An investigation into the dynamics of learners' classroom emotions. *Studies in Second Language Learning and Teaching*, 10(1), 45-65. <http://dx.doi.org/10.14746/ssllt.2020.10.1.3>
- Dinçer, B. H. (2019). *The validity of translation items in English language proficiency exams in Turkey*. (Unpublished master's thesis). Hacettepe University, Ankara: Turkey.



- Douglas, D. (2005). Testing languages for specific purposes. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 857-868). Mahwah, NJ: Lawrence Erlbaum.
- ETS (2019a). *TOEFL iBT test integrated speaking rubrics (Scoring standards)*. Princeton, NJ: Educational Testing Service.
- ETS (2019b). *TOEFL iBT test independent speaking rubrics (Scoring standards)*. Princeton, NJ: Educational Testing Service.
- Farhady, H. (1983). New directions for ESL proficiency testing. In J. Oller (Ed.), *Issues in language testing* (pp. 253-69). Rowley, MA: Newbury House.
- Fulcher, G. (2003). *Testing second language speaking*. Harlow: Pearson Longman.
- Golato, A. (2003). Studying compliment responses: A comparison of DCT's and naturally occurring talk. *Applied Linguistics*, 24, 90-121.
- Grabowski, K. (2009). *Investigating the construct validity of a test designed to measure grammatical and pragmatic knowledge in the context of speaking*. (Unpublished doctoral dissertation). Teachers College, Columbia University: New York.
- Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (pp. 41-58). New York: Academic Press.
- Gumperz, J. (1982). *Discourse strategies*. Cambridge, UK: Cambridge University Press.
- Gür, Ö. (2012). Testing, assessment and KPDS (foreign language proficiency examination for state employees)—What does this exam assess? *Sakarya University Journal of Education*, 2(1), 23–32.
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: Edward Arnold.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Halliday, M. A. K., & Hasan, R. (1989). *Language, context, and text: Aspects of language in a socio-semiotic perspective*. Oxford, UK: Oxford University Press.
- Hudson, T., Detmer, E. & Brown, J. D. (1992). *A framework for testing cross-cultural pragmatics*. Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawaii.
- Hudson, T., Detmer, E. & Brown, J. D. (1995). *Developing prototypic measures of crosscultural pragmatics*. Technical Report 7. Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawaii.
- Hutchby, I. & Wooffitt, R. (1998). *Conversation analysis: Principles, practice and applications*. Cambridge: Polity Press.
- Hymes, D. (1967). Models of the interaction of language and social setting. *Journal of Social Issues*, 23(2), 8-38.
- Hymes, D. (1974). Ways of speaking. In R. Bauman & J. Sherzer (Eds.), *Explorations in the ethnography of speaking* (pp. 433-451). Cambridge, UK: Cambridge University Press.
- Jaszczolt, M. (2002). *Semantics and pragmatics: Meaning in language and discourse*. London: Longman.
- Johnson, M. (2001). *The art of non-conversation*. New Haven, CT: Yale University Press.
- Jones, R. (1985). Second language performance testing: An overview. In P. Hauptman, R. LeBlanc, & M. Wesche (Eds.), *Second language performance testing* (pp. 15-24). Ottawa: University of Ottawa Press.
- Kamaşak, R., Sahan, K., & Rose, H. (2021). Academic language-related challenges at an English-medium university. *Journal of English for Academic Purposes*, 49, 100945. <https://doi.org/10.1016/j.jeap.2020.100945>
- Kasper, G. & Rose, K. (2002). *Pragmatic development in a second language*. Maiden, MA: Blackwell.

- Kıray, G. (2015). Macro-structure analysis of reading comprehension paragraphs of KPDS and YDS exams within years 2003-2013. *Hasan Ali Yucel Journal of Education*, 12(1), 219–233.
- Korsko, P. (2004). *The narrative shape of two-party complaints in Portuguese: A discourse analytic study*. Unpublished doctoral dissertation, Teachers College, Columbia University, New York, NY.
- Kutlu, M. O., Demirogları, G., & Demirogları, B. (2020). Washback effect of higher education institution exam-foreign language test on university candidates' attitudes. *African Educational Research Journal*, 8(2), S121-S130.
- Külekcı, E. (2016). A concise analysis of the Foreign Language Examination (YDS) in Turkey and its possible washback effects. *International Online Journal of Education and Teaching (IOJET)*, 3(4). 303-315. <http://iojet.org/index.php/IOJET/article/view/141/143>
- Lado, R. (1961). *Language testing*. New York: McGraw-Hill.
- Lazarton, A. (1992). The structural organization of a language interview: A conversational analysis perspective. *System*, 20, 373-86.
- Leech, G. (1983). *Principles of pragmatics*. London: Longman.
- Liu, J. (2006). *Measuring interlanguage pragmatic knowledge of EFL learners*. Frankfurt am Main: Peter Lang.
- Locher, M. A., & Sage L. G. (2010). Introduction to interpersonal pragmatics. In M. A. Locher & L. Sage (Eds.), *Interpersonal pragmatics* (pp. 1-13). Berlin, New York: De Gruyter Mouton. doi:10.1515/9783110214338.0.1
- McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Maiden, MA: Blackwell.
- Ochs, E. (1996). Linguistic resources for socializing humanity. In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 407-437). Cambridge, UK: Cambridge University Press.
- Oller, J. (1979). *Language tests at schools*. London: Longman.
- OSYM-Öğrenci Seçme ve Yerleştirme Merkezi (2016). YDS kılavuzu. Retrieved on Monday, March 02, 2020 from <http://dokuman.osym.gov.tr/pdfdokuman/2016/GENEL/EsdegerlikYonergesi26022016.pdf>.
- OSYM (2015). *2015 Yabancı dil bilgisi seviye tespit sınavı (YDS)*. Ankara: OSYM.
- OSYM (2016). *2016 Yabancı dil bilgisi seviye tespit sınavı (YDS)*. Ankara: OSYM.
- OSYM (2017). *2017 Yabancı dil bilgisi seviye tespit sınavı (YDS)*. Ankara: OSYM.
- OSYM (2018). *2018 Yabancı dil bilgisi seviye tespit sınavı (YDS)*. Ankara: OSYM.
- Özmen, K. S. (2011). Washback effects of the inter-university foreign language examination on foreign language competences of candidate academics. *Novitas-ROYAL*, 5(2), 215-228.
- Özer, Z., & Altay, İ.F. (2021). Examining the level of enjoyment and anxiety among Turkish EFL students. *Journal of Language and Linguistic Studies*, 17(Special Issue), 663-671.
- Polat, M. (2019). Washback effect of Turkish foreign language proficiency exam YDS: A scale development study. *Journal of Language and Linguistic Studies*, 15(4), 1252-1268.
- Polat, M. (2020). Washback effects of YDS and YOKDİL on foreign language teaching. *Kastamonu Education Journal*, 28(5), 2188-2202. doi: 10.24106/kefdergi.709027
- Purpura, J. (2004). *Assessing grammar*. Cambridge, UK: Cambridge University Press.

- Purpura, J. (2014). Cognition and language assessment. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1452-1476). Oxford, UK: Wiley-Blackwell.
- Purpura, J. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100, 190-208.
- Purpura, J. (2017). Assessing meaning. In E. Shohamy, L. G. Or, & S. May (Eds.), *Language testing and assessment, encyclopedia of language and education* (3rd ed.) (pp. 33-61). Springer International Publishing AG.
- Roever, C. (2006). Validation of a web-based test of ESL pragmalinguistics. *Language Testing*, 23, 229-256.
- Roever, C. (2018). *Testing of L2 pragmatics and possibilities for measuring implicit knowledge* (Unpublished manuscript). Department of Linguistics & Applied Linguistics, The University of Melbourne, Australia.
- Sartepeci, M., Duran, A., & Ermiş, U. F. (2019). A new trend in preparing for foreign language exam (YDS) in Turkey: Case of WhatsApp in mobile learning. *Education and Information Technologies*, 24(5), 2677-2699. <https://doi.org/10.1007/s10639-019-09893-4>
- Suesattayawong, O. (2006). Exploring trouble in speaking English in Rajvithi hospital. (Unpublished doctoral dissertation). Srinakharinwirot University, Bangkok: Thailand.
- Schegloff, E. (1987). Between micro and macro: Contexts and other connections. In J. Alexander, B. Giesen, R. Munch, & N. Smelser (Eds.), *The micro-macro link* (pp. 207-234). Berkeley, CA: University of California Press.
- Searle, J. (1969). *The philosophy of language*. New York: Oxford University Press.
- Spolsky, B. (1968). Language testing—the problem of validation. *TESOL Quarterly*, (2)2, 88-94.
- Tadayon, F., & Ravand, H. (2016). Using grounded theory to validate Bachman and Palmer's (1996) strategic competence in EFL graph-writing. *Lang Test Asia* 6, 8. <https://doi.org/10.1186/s40468-016-0031-y>
- Taguchi, N. (2018). Data collection and analysis in developmental L2 pragmatics research: Discourse completion test, role-play, and naturalistic recording. In A. Gudmestad & A. Edmonds (Eds.), *Critical reflections on data in second language acquisition* (pp. 7–32). Amsterdam: John Benjamins.
- Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics*, 4, 91-112.
- Thomas, J. (1995). *Meaning in interaction: An introduction to pragmatics*. London: Longman.
- Timpe-Laughlin, V., Green, A., & Oh, S. (2021). Raising pragmatic awareness: A think aloud study. *System*, 98, 102470. <https://doi.org/10.1016/j.system.2021.102470>
- Yamashita, S. (1996). *Six measures of JSL pragmatics*. Technical Report 14. Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawaii.
- Yamashita, S. (2001). Using pictures for research in pragmatics. Eliciting pragmatic strategies by picture response tasks. In T. Hudson & J. D. Brown (Eds.), *A focus on language test development* (pp. 35-56). Honolulu, HI: University of Hawaii, National Foreign Language Resource Center.
- Youn, S. J. & Bi, N. Z. (2019). Investigating test-takers' strategy use in task-based L2 pragmatic speaking assessment. *Intercultural Pragmatics*, 16, 185–218.
- Yule, G. (1996). *Pragmatics*. Oxford, UK: Oxford University Press.
- Zeybek, G. (2015). *Foreign Language Examination (FLE) test takers' foreign language learning attitude, motivation and test anxiety in Turkey*. (Unpublished master's thesis). Süleyman Demirel University, Isparta, Turkey.