

## Article

# Standardization of the Gender Assignment and Agreement Assessment in the Greek Language: Preliminary Evidence from Bilingual Greek—Albanian School Age Children

Alexandra Prentza <sup>1</sup>, Maria Kaltsa <sup>2</sup>, Dionysios Tafiadis <sup>3,\*</sup>  and Ianthi Maria Tsimpli <sup>4</sup>

<sup>1</sup> Department of Linguistics, University of Ioannina, 45110 Ioannina, Greece; aprentza@uoi.gr

<sup>2</sup> Department of Theoretical and Applied Linguistics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; mkaltsa@enl.auth.gr

<sup>3</sup> Department of Speech and Language Therapy, University of Ioannina, 45110 Ioannina, Greece

<sup>4</sup> Faculty of Modern and Medieval Languages and Linguistics, University of Cambridge, Cambridge CB2 1TN, UK; imt20@cam.ac.uk

\* Correspondence: tafiadis@uoi.gr

**Abstract:** (1) Background: Given the scarcity of data on the psychometric evaluation of measures used with typically developing Greek-speaking bilinguals, this study aims to present preliminary evidence for the standardization and the psychometric evaluation of a gender assignment and agreement assessment designed for the Greek language (henceforth, GAAGL Assessment) employing real and pseudo-words. This is the first study to standardize the GAAGL Assessment and to explore its discriminatory ability with typical populations. (2) Methods: The assessment was designed as part of the BALED project which targeted language skills in bilingual children for whom one language is Greek. For the psychometric evaluation and the exploration of the test's discriminatory ability we ran a Cronbach's alpha analysis and a Youden Receiver Operating Characteristics (ROC) analysis across the domains and sub scores of the test. Our sample consisted of 53 typically developing Greek monolingual children and 57 age-matched Greek–Albanian-speaking bilinguals. (3) Results: The results showed: (a) a high internal consistency for the GAAGL Assessment across its four main tasks and (b) excellent discriminatory ability, since statistically significant positive discrimination was detected between monolingual and bilingual children in the tasks and scores of the GAAGL Assessment.

**Keywords:** standardization; assessment; Greek gender; typical language development; monolingual; bilingual



**Citation:** Prentza, Alexandra, Maria Kaltsa, Dionysios Tafiadis, and Ianthi Maria Tsimpli. 2023. Standardization of the Gender Assignment and Agreement Assessment in the Greek Language: Preliminary Evidence from Bilingual Greek—Albanian School Age Children. *Languages* 8: 136. <https://doi.org/10.3390/languages8020136>

Academic Editor: Elena Babatsouli

Received: 21 March 2023

Revised: 12 May 2023

Accepted: 20 May 2023

Published: 26 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The main objective of the present study is to present preliminary evidence for the standardization of a battery of gender assignment and agreement assessment tasks for the Greek language (GAAGL Assessment hereafter) considering both real words and pseudo-words. For the assessment of gender assignment, participants are expected to provide an agreeing determiner, while for gender agreement, participants are expected to provide an agreeing adjective predicate. This yielded four main tasks: the determiner task with real words (det\_real), the determiner task with pseudo-words (det\_pseudo), the adjective task with real words (adj\_real) and the adjective task with pseudo-words (adj\_pseudo). The GAAGL Assessment was designed for the purposes of the BALED project (MIS377313), which targeted language skills in bilingual children for whom one language is Greek, and the other is either Albanian, English or German. In the present study we focus on the Greek–Albanian language pair.

Gender is often used as a measure of grammatical development in bilinguals due to the fact that it provides an insight in bilingual grammatical development where one language grammaticalizes a grammatical feature, while the other either does not or grammaticalizes

it differently (for gender studies see [Bianchi 2013](#) on Italian/German, [Balam et al. 2021](#) on Spanish/English, [Mitrofanova et al. 2022](#) on German/Russian, [Kupisch et al. 2013](#) on German/French, [Pérez-Tattam et al. 2019](#) on Basque/Spanish, as well as the studies of [Alexiadou et al. 2021](#) on English/Greek, [Chondrogianni 2007, 2008](#) on Turkish/Greek, [Kaltsa et al. 2019](#) on English/Greek and German/Greek, [Konta 2013](#) on Turkish/Greek and [Prentza et al. 2019](#) on Albanian/Greek). Hence, in this study, stepping away from merely identifying the cross-group differences with respect to gender, we opted to explore the validity and reliability of an assessment that exploits gender and touches both on the lexicon (real words) as well as on the grammatical system (pseudo-words) in two syntactic operations, namely gender assignment and agreement, exploring, therefore, the interaction among those parameters in language development. Additionally, we consider the discriminatory ability of this test as measured via the estimated cut-off scores between monolingual and bilingual children as a first step towards the standardization of the tool.

### 1.1. Gender Distinctions in Greek and Albanian

Cross-linguistically, gender encoding instantiates a tripartite distinction, masculine, feminine and neuter. (However, there are languages such as Swahili which show four gender values.) According to [Corbett's \(1991\)](#) classification, gender can be predicted on the basis of phonological, morphological and semantic rules, while the hierarchy of those rules appears to be language-specific. In languages that parameterize gender, grammatical gender is identified as a lexical classification feature (cf. [Chomsky 1995](#)) and gender agreement among nominal elements is a syntactic computation that entails a dependency of a gender feature among nominal elements; specifically, an uninterpretable/unvalued one on the determiner (D) or adjective (Adj) and its valued counterpart on the noun (N) ([Carstens 2010](#); [Chomsky 2001](#); [Bošković 2009](#)).

Greek is a grammatical gender language where nouns are assigned a valued, uninterpretable gender feature in the lexicon and the gender agreement among the nominal elements is regulated by the unvalued/uninterpretable gender feature (either on D or Adj) targeting the valued gender feature on the noun ([Tsimpli and Hulk 2013](#)). In other words, for any noun that is a lexical entry in the mental lexicon of a speaker, that is for any real word, gender is deterministic, while for any novel noun, gender is assigned via the utilization of other cues, such as morpho-phonological information on noun endings. In Greek, grammatical gender bears a tripartite distinction, i.e., masculine, feminine and neuter and is phonetically realized through morphological marking on all nominal elements, that is determiners, adjectives, numerals, pronouns and quantifiers. According to [Ralli \(2002\)](#), the gender feature on the noun stem is reflected via the suffix alone (cf. [Hawkins and Tsimpli 2009](#); [Tsimpli 2003](#)). This feature establishes gender agreement and gender assignment on pronominal and nominal forms (see examples (1), (2) and (3) below).

- (1) Masculine (Masc)  
o v'ðinemos 'enðres  
theMASC weakMASC manMASC
- (2) Feminine (Fem)  
i 'omorfi 'kori  
theFEM beautifulFEM daughterFEM
- (3) Neuter (Neut)  
to psi 'lo vu 'no  
theNEUT highNEUT mountainNEUT

Greek nouns fall into various inflectional classes depending on the number of syllables in each number (singular, plural), their suffixes and gender, among other features ([Holton et al. 1997, 2012](#); [Ralli 2002, \[2005\] 2022](#); [Triandaphyllidis 1949](#)). Table 1 provides examples ([Holton et al. 1997](#), p. 70).

**Table 1.** Greek gender marking across inflectional classes (adapted from (Holton et al. 1997, p. 70)).

Masc		Fem		Neut	
-as	pe'teras 'father'	-a	'θelese 'sea'	-o	vi'vlio ' book'
-is	'kleftis 'thief	-i PL: -es	'kori ' daughter'	-i	a'γori 'boy'
-os	aðel'fos 'brother'	-os	'isoðos 'entrance'	-os	'leθos ' error'
-eas	ku'rees 'gents' hairdresser'	-i PL: -is	'poli 'town'	-ma -imo	'meθime 'tutorial' 'γρεpsimo 'writing'

According to Ralli (2002, [2005] 2022), gender is not the sole factor that determines the categorization of noun inflectional classes in Greek, but rather other morphological criteria are employed. Anastasiadi-Symeonidi and Cheila-Markopoulou (2003), suggesting that both morphological and semantic criteria shape the prototypicality of suffixes, argued that -os, -as and -is are closely related to masculine gender, -a, -i and -u to feminine, while -o and -i to neuter. Their suggestion is that prototypicality is associated with frequency of occurrence, an assumption which is supported by psycholinguistic data, since monolingual children and L2 learners learn prototypical nouns first. Additionally, according to Tsimpli and Hulk (2013), neuter in Greek appears to function as a linguistic and learner default and, as such, input-related properties such as the level of transparency, the frequency of suffixes per gender and the consistency in gender marking, along with the learner's 'strategy' affect acquisition and use of the neuter value in Greek. Anagnostopoulou (2017) adds that the neuter is the linguistic and learner default as far as (-animate) nouns are concerned, and while masculine is considered to be the default for (+human), the default of (-human) (+animate) nouns (i.e., animals) remains unclear. Based on the above, for the design of the GAAGL Assessment we employed the analyses on the predictive values of noun endings in Greek and chose noun endings that are strongly associated with gender values (Table 1, as well as Section 5). Additionally, the employment of pseudo-words along with real words aided to the exploitation of gender attribution, since it eliminates any lexical skills interference; therefore, both real words and pseudo-words across gender values (masculine, feminine and neuter) were used in the GAAGL Assessment that we evaluate in this paper.

The bilingual population we tested in the current study are children speaking both Greek and Albanian. Albanian is also a grammatical gender language which has a productive two-way gender distinction, masculine and feminine, marked on determiners and nouns (see Kurani and Trifoni 2014; Millaku and Topanica-Millaku 2021; Revithiadou and Spyropoulos 2013). According to Millaku and Topanica-Millaku (2021) masculine nouns in Albanian appear with the endings -u, -un, -ut, -i, -in, -it, for example "libri" (the book) while feminine nouns appear with the endings -a, -e, -ja, -je, -në, -s, -së for example "motra" (the sister). The use of the neuter is restricted to cases such as "the NEUT walking" (Agalliu 2002). According to Giurgea (2014), gender agreement in Albanian is dependent on number, since the language exhibits a productive class of nouns ("ambigeneric") which show masculine agreement in the singular and feminine agreement in the plural (e.g., the noun "tregim" (story), tregim (sg.masc.)/tregime (pl.fem.)). In light of the above, it could be argued that the feminine is the default form in the plural; however, for the singular, the default form cannot be easily identified. Due to our bilingual participants' experience with two languages that grammaticalize gender diversely, we expect them to be sensitive to the linguistic cues available for the assignment of the gender value.

### 1.2. Monolingual and Bilingual Greek Speakers and the Acquisition of Gender

Earlier research suggests that by the age of 3;6, monolingual Greek children acquire gender (Marinis 2003; Mastropavlou 2006; Stephany 1997; Stephany and Christofidou 2008). As analysed in Section 2, while neuter value is considered the default, masculine and femi-

nine are marked, from the early stages of development (Mastropavlou 2006; Stephany 1997). Additionally, Stephany and Christofidou (2008) note that neuter nouns appear to have high frequency both in child-directed and in the child's early speech production, while feminine and masculine nouns appear to be less frequent in both language measurements. Hence, it is no surprise that neuter appears to be acquired easily in both L1 and L2 Greek.

Turning to the predictive value of noun suffixes, suffixes with high predictive value can decidedly predict the gender value of nominal elements, whereas with suffixes with low predictive value, it is uncertain which gender value they carry. Elicitation data on Greek pseudo-nouns has shown that the monolinguals are sensitive to noun suffixes, yet a certain length of exposure is required for predictive values to develop. Specifically, Varlokosta (2011) tested the production of D-N and D-Adj-N contexts in five-year-old monolingual Greek-speaking children and found that accuracy rises in cases of suffixes with high predictive values (specifically, -as and -is for masculine nouns, -o for neuter nouns) and it drops in suffixes with low predictive value (such as -i). Mastropavlou and Tsimpli (2011) explored the differences between oral elicitation of novel noun DPs and elicitation with orthographic cues in adult monolingual speakers of Greek and reported that the predictive value of suffixes was high when orthographic cues which resolved ambiguity were provided, whereas when no orthographic cues were provided, the predictive value of the suffix -i, which can be associated either with neuter or with feminine nouns, dropped significantly.

Regarding the use of gender by Greek-speaking populations who are raised as bilinguals, Chondrogianni (2007, 2008) tested seven to twelve-year-old Turkish-Greek-speaking sequential bilingual children. Her data shows that at the early stages of gender acquisition there are some difficulties but, as proficiency increases, the use of gender marking becomes native-like. Additionally, Chondrogianni notes that Adj-N gender agreement errors appear to be more frequent than D-N gender assignment errors, in a follow-up study of the same type of bilingual children. In the same language pair paradigm, Konta (2013, 2019) explored gender assignment and agreement in Greek with Greek-Turkish bilingual adolescents and showed that morphophonological cues are crucial for D-N contexts, but not for gender agreement in Adj-N contexts which were more challenging when relying on animacy distinctions. Prentza et al. (2019) investigated the acquisition of gender in Albanian-Greek-speaking bilingual children and found that while lexical knowledge and quality of input can assist the acquisition of Greek gender, gender agreement is overall delayed and gender attribution scores remained lower in the pseudo-words compared to real words. These findings were explained as follows: (a) the assignment-agreement differences were attributed to the structural complexity of Adj-N agreement in a predicate construction compared to D-N agreement within the DP, leading to more errors (see also Tsimpli 2003 on adult L2ers) and (b) for real-pseudo differences, the asymmetry is related to the different mechanisms involved, namely, stored lexical knowledge for the real word (Ralli 2002, [2005] 2022), as opposed to mere exploitation of noun endings for pseudo-words (Tsimpli and Hulk 2013). Meanwhile, Kaltsa et al. (2019) explored gender assignment and agreement by comparing data from bilingual children speaking Greek and English, as well as Greek and German and found: (a) evidence of positive language transfer when both languages grammaticalize gender while the neuter value was used as a default; (b) that gender agreement was more demanding; and (c) that pseudo items were more challenging for bilingual children. These findings were also verified by Kaltsa et al. (2020), who compared Albanian-Greek to English-Greek bilinguals on gender attribution with overall lower performance in gender agreement, especially in the case of pseudo-words and better performance in gender attribution to masculine and neuter nouns than to feminine ones. Additionally, Babatsouli and Nicoladis (2019), in a longitudinal study, examined the acquisition of English possessives by a bilingual Greek-English-speaking child and found that the acquisition of possessives in the dominant language, i.e., Greek, facilitates the acquisition of possessives (that encompass gender) in the weaker language, i.e., English. Recently, Alexiadou et al. (2021) tested gender agreement in adolescent heritage speakers

of Greek whose other language is English and reported that most gender mismatches are identified in noun phrase internal agreement contexts and verified that neuter is used as the learner default for L2 Greek.

### 1.3. Standardization of Language Tasks with Typically Developing Bilingual Populations

Generally, bilingual children are reported to score below monolingual children in grammar tasks, such as the gender tasks reported in Section 3. However, as [Plante and Vance \(1994\)](#) pointed out, between-group differences alone cannot be used to characterize a task as an accurate tool for discriminating populations, in this case monolinguals from bilinguals. Apart from its discriminatory ability, any psychometric evaluation of a test also considers the validity and the reliability of the instrument. Validity evaluations examine whether a concept or ability is accurately measured by the instrument, while reliability evaluations examine whether repeated applications of the instrument would provide the same results ([Heale and Twycross 2015](#)). Given that monolingual and bilingual language development may present differences in certain language areas, with typically developing bilingual children showing delays compared to monolingual peers (see [Tsimpli 2014](#) for an overview), bilingual children run the risk of being inaccurately identified as having a language impairment. Therefore, it is crucial that researchers and educators know the range within which typically developing bilingual children fall with respect to various grammatical measures, especially ones that are frequently used with specific populations, such as gender, both for the accurate assessment of language skills in bilinguals as well as for the accurate diagnosis of impairment in some cases (see also [Marinis et al. 2017](#); [Prentza et al. 2022](#)).

For these reasons, a growing body of normative data for typically developing bilingual children is now available. However, the emphasis remains on bilingual populations for whom one language is English, while the language domain that has been extensively examined is the lexicon; some examples are the Peabody Picture Vocabulary Test (PPVT-III; [Dunn and Dunn 1997](#)), which examines three groups of bilinguals for whom one language is American English, the British Picture Vocabulary Scale ([Dunn et al. 1997](#)), which presents normative data for Welsh–English bilingual children ([Mueller Gathercole et al. 2008](#)), as well as the Boston Naming Test ([Kohnert et al. 1998](#)) with preliminary normative data for Spanish–English-speaking children. Concerning tools that explore other areas of language ability, the cases remain limited; Rosselli and colleagues, in their study ([Rosselli et al. 2010](#)), presented preliminary normative data for bilingual children of 5–14 years of age who speak Spanish and English using a language and neuropsychological battery, while the Bilingual English Language Assessment (BESA) tool ([Peña et al. 2018](#)) was designed to examine morpho-syntax, phonology, semantics and pragmatics in both languages of the bilinguals, namely, Spanish and English, and offered norms for all of the above domains.

Because of the emphasis placed in English-speaking populations, along with the fact that Greek is a less frequently used language, currently, there are scarce data on validated grammar tools used with Greek-speaking bilinguals. Only two previous studies presented relevant data, yet for overall grammar measures such as Sentence Repetition Tasks (SRTs) such as Sentence Repetition Tasks (SRTs). Specifically, [Theodorou et al. \(2017\)](#) evaluated the ability of an SRT designed in Cypriot Greek to discriminate between typically developing bidialectal Cypriot-Greek children and children with a developmental language disorder, with the data showing that the task can accurately discriminate between the two populations. Recently, [Prentza et al. \(2022\)](#) explored the psychometric features, namely, the validity, reliability and discriminatory ability of the LITMUS Greek SRT with typically developing bilingual and monolingual children. The results showed that the instrument scored very high in the validity and reliability measures, while accurate discrimination of the two groups was achieved, since different cut-off scores were calculated for monolingual and bilingual children. Note crucially that the study of [Prentza et al. \(2022\)](#) uses the same population pool as the present study. We think that this line of investigation is interesting, since keeping the population constant while alternating the tasks, one for



overall grammar ability and one targeting a specific grammatical feature, could reveal the psychometric characteristics of the tasks more easily without any confounding results as an effect of population characteristics or size. Moreover, given that gender is a sensitive grammar task, in the sense that its administration tends to reveal differences between monolingual and bilingual populations, as also shown in Section 3, the existence of a standardized grammatical gender measure for the Greek language could help researchers and educators discriminate between different types of Greek-speaking populations via the administration of a specific grammar test. Finally, since Albanian, the other language of the bilinguals, grammaticalizes gender, this would yield interesting results regarding the syntactic operations required in the different components of the GAAGL Assessment.

Considering the above, this paper aims to evaluate the psychometric properties of the GAAGL Assessment; specifically, it aims to examine the validity and reliability of the test as well as to test its discriminatory ability with monolingual and bilingual typically developing Greek-speaking children. This study is novel since it attempts to standardize a grammar-specific language tool with typically developing Greek-speaking monolingual and bilingual children, something not attempted before. The selection of a gender task adds to the value of the study given that (a) gender tasks are frequently used with bilingual children, and (b) bilingual Greek-speaking children tend to score lower than monolinguals in such measures. Therefore, researchers and educators need a baseline of typical bilingual language ability for the assessment of Greek-speaking bilinguals and before diagnosing any impairment in these populations.

## 2. Materials and Methods

### 2.1. The Participants of the Study

In total, 53 Greek monolingual children (mean age: 9.87 years ( $SD = 1.3$ ), age range: 8–12), and 57 age-matched bilinguals who were speakers of both Greek (the school language) and Albanian (the family language) (mean age: 9.9 years ( $SD = 1.09$ ), age range: 8–12) participated in the study. Of the participants, 63 were boys and 47 were girls (monolinguals: 28 boys, 25 girls, bilinguals: 27 boys, 30 girls). Children attended Greek state schools in Epirus, Northwest Greece, where Greek is exclusively used for instruction. The National Institute of Education approved the research and access to school was given to the research team. As shown by their school records, child participants did not have learning difficulties or other disabilities. In the cognitive tasks which functioned as screening tests, monolinguals and bilinguals did not differ and both groups achieved age-appropriate scores. Specifically, in the Raven's Coloured Progressive Matrices Task (Raven et al. 1998) monolinguals scored 82% (raw score: 30/36  $SD = 7.2$ ), while bilinguals scored 75.5% (raw score: 27/36,  $SD = 9.5$ ) with no significant difference found between the groups ( $t(108) = 1.7$ ,  $p > 0.05$ ). In the Digit Backwards Span Task (Alloway 2007), the monolinguals scored 70% (raw score: 24/36,  $SD = 9.4$ ) while the bilinguals 73%: (raw score: 26/36,  $SD = 11.3$ ), with no significant differences between the groups ( $t(108) = 1.4$ ,  $p > 0.05$ ). Additionally, an extensive background questionnaire (Mattheoudakis et al. 2016) administered to both parents and children detailed the profile of bilinguals. Particularly, children were early sequential L1 Albanian L2 Greek bilinguals, since they were exposed to Greek at around age 2 (mean age of onset of exposure: 1;9 ( $SD = 0.9$ )) and they had received extensive exposure to the second language, since the length of exposure to the language was calculated at 7;2 ( $SD = 1.2$ ). The frequency of the use of the two languages at the preschool age was comparable, as was the current use of the languages (no significant differences detected,  $ps > 0.05$ ); however, the trends were reversed, with children using the family language more frequently in the preschool years, while Greek was more frequently used in the school years (Language use at the preschool years: Alb = 55% ( $SD = 9.8$ )/Gr = 45% ( $SD = 10.5$ ). Language use in the school years: Alb = 46% ( $SD = 9.4$ )/Gr = 54% ( $SD = 8.7$ )). Additionally, the children reported that they prefer to use Greek for daily writing/reading activities (i.e., reading magazines, writing shopping lists etc.), since most of them did not have literacy skills in Albanian (Gr = 63% ( $SD = 3.5$ ), Alb = 5% ( $SD = 2.3$ ),  $p < 0.01$ ). Finally, bilinguals fared

significantly lower than monolingual peers in a standardized Greek expressive vocabulary test administered (Vogindroukas et al. 2009) (monolinguals: 42.1/50 (84.2%, SD = 4.5), bilinguals: 36.8/50 (73.6%, SD = 5.6), ( $t(108) = 2.7$ ,  $p < 0.05$ ).

## 2.2. The GAAGL Assessment

The GAAGL Assessment comprised four main tasks: two gender assignment tasks (real and pseudo-nouns) and two gender agreement tasks (real and pseudo-nouns) in Greek. In the gender assignment tasks (or gender determiner tasks), participants had to orally produce a definite or indefinite article (D–N) that agrees in gender with the real word or pseudo-word provided to them (i.e., *det\_real*, *det\_pseudo*), while in the gender agreement tasks, they had to orally produce a predicate adjective that agrees with the real word and pseudo-word presented (i.e., *adj\_real*, *adj\_pseudo*). All test items appeared in the nominative singular form. The real noun task involved 54 items, all items were non-human and were either animals (+animate) (N: 24) or everyday objects (-animate) (N: 30)<sup>1</sup>. Additionally, with regard to the selection of real nouns, we opted for items with similar frequency and level of familiarity; specifically, we considered the Hellenic National Corpus data during the experimental built-up and the familiarity of the items was evaluated during the piloting phase on the basis of data provided by monolingual Greek-speaking children attending the 3rd grade. During the task administration, first, participants were familiarized with the test items, which were presented to them both orally (pre-recorded female voice in Standard Greek) and visually (presentation of pictures depicting the items). This step was included to minimize frequency-driven lexical access effects in the dataset. Note that the word was also presented to the children in a written form on the computer screen, providing also orthographic cues which could assist them with ambiguous suffixes (i.e., -i is spelled as -η in feminine nouns while as -ι in neuter nouns). After three practice items, children proceeded with the actual test. On the computer screen, a picture of a table with different items on top of it appeared while the child had to name the animal or the object that appeared on the table. See example (4) where the item that appears on the table is a cat:

- (4) Question: ti 'inɛ'pɛno sto trɛ'pɛzi?  
 what be3SING on the table  
 'What appears on the table?'  
 Expected response: miɛ/i 'ɣɛtɛ  
 aFEM/theFEM catFEM  
 'A/the cat.'

For each of the 3 Greek gender values, 18 items were designed; for the masculine and neuter value, 6 items for each of the 3 endings were designed, (i.e., masculine: -os, -as, -is. neuter: -i, -o and -ma for neuter), and in the feminine value, 9 items for each of the 2 endings (-a, -i) were designed (see also Table 1). In the pseudo-word category, 56 test items which consisted of 3 syllables were designed and were accented in the penultimate syllable. The suffixes used by gender value were the same as the affixes in the real words; however, this time 8 items were employed for each ending (-is, -as, -os, -a, -i (in feminine nouns spelled as η), -o, -ma). With regard to the pseudo-item construction, we built items that would not be easily relatable to real words; in this attempt we altered both a phoneme at the initial consonantal cluster along with other changes such as accent and other phonemes so that children would not relate them to existing words—examples of such items include “spi'xɛni” and “kro'sopo”. Final items selected were also piloted in monolingual Greek-speaking children. The visual stimuli developed for the elicitation of these items were non-human, inanimate objects, so as to eliminate the possibility of such values affecting the performance of our participants. During the administration of the task, the participant viewed the novel object corresponding to the pseudo-word, while at the same time listened to the cue and read its linguistic form (oral and graphemic presentation). Next, s/he was presented with the novel object on a table and was asked the question ‘What is on the

table?’ Example (5) provides a model response when the participant is presented with a pseudo-noun with the masculine suffix -as:

- (5) 'enɐs/o vrɐ'tirɐs  
aMASC/theMASC vratirasMASC  
“a/the vratiras”

In the gender agreement tasks (or gender adjective tasks) children had to provide an agreeing predicate adjective. As in the gender attribution task, the real noun task comprised 54 test items, while the pseudo-noun task had 56. The suffixes employed per gender value as well as the training part were kept constant. However, in the adjective tasks, two different pictures of each item (real or pseudo-noun) were presented in each slide depicting differences in size/color. With regard to the proportion of size and color adjectives elicited, there was an equal distribution between the two types across items and the order of presentation of items was randomized. An arrow also appeared which drew attention to one of the two items while children were presented with the following question: ‘What kind of X (item) is it?’:

- |        |  |                        |
|--------|--|------------------------|
| (6) a. | ti ku'tɛli 'inɛ? mɛ'ɣɛlo<br>ti spoon NEUT be3SING? bigNEUT<br>‘What kind of a spoon is it? Big’  | Real item—real noun    |
| b.     | ti vrɐ'tirɐs 'inɛ? 'kocinos<br>ti vratirasMASC be? redMASC<br>‘What kind of vratiras is it? Red’ | Novel item—pseudo-noun |

### 2.3. Scoring Procedure and Statistical Analysis

Responses were transcribed and participant accuracy scores were calculated. Each correct response was assigned a score of 1, while no score was given for erroneous responses. Correct responses were determined on the basis of the following: (a) in the assignment (determiner) tasks (real or pseudo), participants had to use a determiner that agrees in gender with the noun presented and (b) in the agreement (adjective) tasks (real or pseudo), participants had to use any adjective that agrees in gender with the noun presented. The Kolmogorov–Smirnov and Shapiro–Wilk tests were performed to examine the distribution of the variables. All normally distributed experimental variables were expressed through mean scores (M) and standard deviation scores (SD). The comparison between monolingual and bilingual groups with regard to all gender tasks was calculated using Student’s *t*-test.

Additionally, a Receiver Operating Characteristics (ROC) analysis was run to evaluate the test’s discriminatory ability with monolingual and bilingual children (Eckes 2017; Fan et al. 2006; Huang and Konold 2014), while for the computation of the cut-off scores, the Youden analysis (Youden 1950) was applied, since it is a robust and effective analysis for the calculation of cut-off scores (Fluss et al. 2005; Martínez-Cambor and Pardo-Fernández 2019; Mizuta et al. 2022; Perkins and Schisterman 2005). ROC analysis sensitivity and specificity values which range between 80% and 89% are considered fair, while rates above 90% are considered good (Plante and Vance 1994; Rhodes et al. 2019). The area under the curve (AUC) measures the test’s discriminatory ability, with the maximum value for AUC being 1 and the minimum 0. According to Chondrogianni and John (2018), AUC scores can be interpreted as follows: an area of 1 shows a perfect test; an area of 0.90–1 indicates excellent accuracy; an area of 0.80–0.90 shows good accuracy; an area of 0.70–0.80 shows fair accuracy; an area of 0.60–0.70 reveals poor accuracy. An area below 0.60 denotes a test that is neither valid nor useful. In terms of figure interpretation, the straight diagonal line that appears represents a baseline of no discriminatory value (50% specificity, 50% sensitivity), while the closeness of the apex of the curve toward the upper left corner is directly proportional to the discriminatory ability of the test (Fan et al. 2006). In our study, three separate ROC analyses were conducted in order to examine the test’s discriminatory ability as follows: (a) in the four main tasks of the Assessment, namely real and pseudo-noun assignment and agreement, (b) in the three gender type scores, namely, masculine,



feminine and neuter and (c) in the scores of gender types X noun type, namely real and pseudo masculine, real and pseudo feminine and real and pseudo neuter.

Finally, the internal consistency of the GAAGL Assessment was evaluated via the computation of a Cronbach's alpha coefficient and a split-half reliability coefficient measure (Krzanowski and Woods 1984). In the Cronbach's internal analysis, the measure of consistency examines whether and to what extent all the items in a test measure the same concept, while the measure of reliability evaluates the amount of measurement error in a test. Cronbach's alpha analysis returns a number between 0 and 1 with the acceptable values appearing between 0.70 and 0.95, while results higher than 0.95 may indicate redundancy among the components of a test (Nunnally and Bernstein 1994; Ursachi et al. 2015). Given that the Cronbach's alpha analysis examines the contribution of each main component to the reliability of a test, we ran the analysis for the main tasks of the GAAGL Assessment, namely, gender assignment in real and pseudo-nouns as well as gender agreement in real and pseudo-nouns.

All reported *p* values were two-tailed with statistical significance set at  $p < 0.05$ . The SPSS statistical software (version 19.0, Armonk, NY, USA) was used for the data analysis.

### 3. Results

#### 3.1. Between-Group Comparisons

Starting with the between-group comparisons of the four tasks of the GAAGL Assessment, namely, gender assignment in real nouns (det\_real), gender assignment in pseudo-nouns (det\_pseudo), gender agreement in real nouns (adj\_real) and gender agreement in pseudo-nouns (adj\_pseudo), the analysis showed that the monolingual group achieved statistically significantly higher scores than the bilingual in all four sub-components. Between-group comparisons and relevant results are presented in Table 2. The maximum score for the determiner tasks is 54, while for the adjective tasks it is 56.

**Table 2.** Between-group Comparisons—GAAGL Assessment Tasks.

	Monolinguals		Bilinguals		<i>t</i> (108)	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
det_real (/54)	53.96	±0.192	50.04	±5.51	5.181	<0.001 *
det_pseud (/56)	50.55	±2.69	43.32	±5.36	8.843	<0.001 *
adj_real (/54)	53.98	±0.137	51.30	±4.35	4.486	<0.001 *
adj_pseud (/56)	51.68	±3.36	36.14	±8.26	12.749	<0.001 *

Notes: det\_real, determiner real words; det\_pseud, determiner pseudoword; adj\_real, adjective real words; adj\_pseud, adjective pseudowords; SD, standard deviation; \*  $p < 0.01$ .

As shown in Table 2, while pseudo-words pose a challenge for all children, they are related with particularly lower scores in the bilingual group.

Next, we proceeded with between-group comparison by gender value scores. The analysis returned that monolinguals outperformed bilinguals in the gender assignment and agreement tasks of the GAAGL Assessment across all three gender values. Table 3 presents the relevant results. The maximum score for masc\_total is 84, while the maximum score for feminine\_total and neuter\_total is 68.

**Table 3.** Between-group comparisons—GAAGL Assessment gender value scores.

	Monolinguals		Bilinguals		<i>t</i> (108)	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Masc_total (/84)	83.94	±0.412	73.58	±8.86	8.501	<0.001 *
Fem_total (/68)	61.66	±4.07	48.94	±8.17	10.210	<0.001 *
Neut_total (/68)	64.21	±2.56)	58.24	±6.53	6.215	<0.001 *

Notes: Masc\_total, masculine total score; Fem\_total, feminine total score; Neut\_total, neutral total score; SD, standard deviation; \*  $p < 0.01$ .

Across groups, these results show that gender assignment in masculine nouns poses the least problems. Additionally, for bilingual children, the lowest scores are attained in feminine nouns, (48.94/68) with the neuter nouns following (58.24/68).

Next, we proceeded with between-group comparison both by noun type (real or pseudo) and by gender value. As in the previous analyses, the monolinguals performed significantly better than the bilinguals in all gender marking types. Results are presented in Table 4. The maximum score for real nouns is 36, the maximum score for masculine pseudo-nouns is 48, while for feminine and neuter nouns, it is 32.

**Table 4.** Between-group comparisons—GAAGL Assessment—noun type and gender value scores.

	Monolinguals		Bilinguals		<i>t</i> (108)	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
masc_real (/36)	36.00	±0.000	33.91	±3.25	4.671	<0.001 *
masc_pseudo (48)	47.94	±0.412	39.66	±6.83	8.801	<0.001 *
fem_real (/36)	36.00	±0.000	33.22	±4.09	4.929	<0.001 *
fem_pseudo (/32)	25.66	±4.07	15.71	±5.23	11.048	<0.001 *
neut_real (/36)	35.94	±0.233	33.84	± 3.58	4.258	<0.001 *
neut_pseudo (/32)	28.26	±2.52	24.40	±3.90	6.110	<0.001 *

Notes: masc\_real, masculine real words; masc\_pseudo, masculine pseudo-words; fem\_real, feminine real words; fem\_pseudo, feminine pseudo-words; neut\_real, neutral real words; neut\_pseudo, neutral pseudo-words; *SD*, standard deviation; \* *p*-level < 0.01.

As these results show, the bilingual children score the lowest in the feminine and neuter pseudo-words, both compared to other word types within their group, as well as when compared with monolingual children.

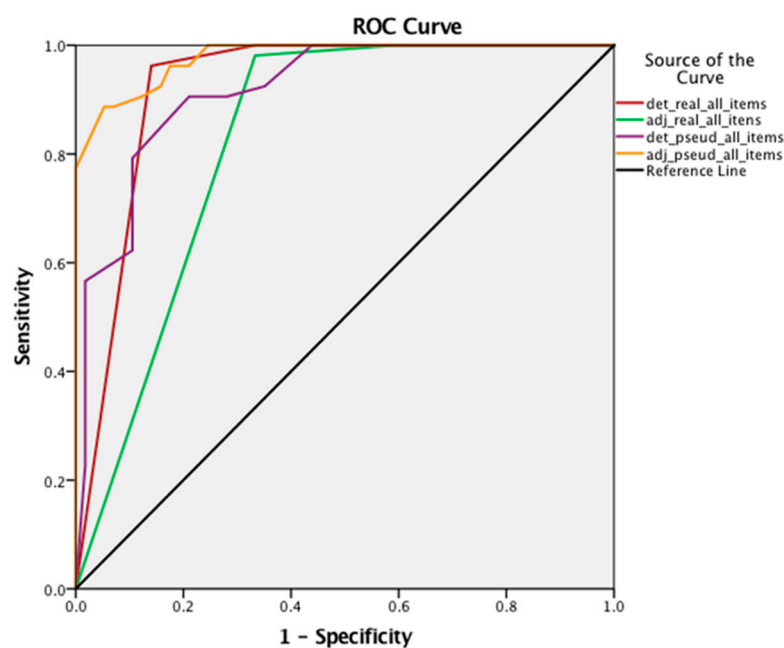
### 3.2. ROC Analyses

However, given that the examination of between-group comparisons cannot reveal the discriminatory ability of a measure, we proceeded with applying an ROC analysis on the GAAGL Assessment. Three separate ROC analyses were conducted in order to examine if there is any statistically significant positive discrimination between the monolingual and the bilingual group as follows: (a) in the four main tasks of the Assessment, (b) in the three gender type scores and (c) in the gender type X noun type scores.

Starting with the first ROC analysis on the GAAGL Assessment's four tasks, the results showed a very high statistically significant positive discrimination between the monolingual and bilingual participants of our study in all four tasks of the Assessment. Relevant data is presented in Figure 1 (ROC curve) and Table 5 (cut-off points, AUC, *p*-value).

The AUC scores were excellent and ranged from 0.828 to 0.978, with the highest scores attained in the adj\_pseudo task followed by the det\_real task. The cut-off scores in the real determiner and real adjective task were calculated quite high, at 53.5/54, while in the pseudo-words, both in the determiner and in the adjective task, they ranged close to 48/56.

Similarly, a separate ROC analysis was conducted to determine if there is any statistically significant positive discrimination between the monolingual and the bilingual groups in the GAAGL Assessment scores by gender value. The results showed a high positive discrimination between the two groups for each of the gender values of the items in the GAAGL Assessment. Figure 2 presents the ROC curve, while Table 6 shows the cut-off scores, the AUC and the *p*-value.

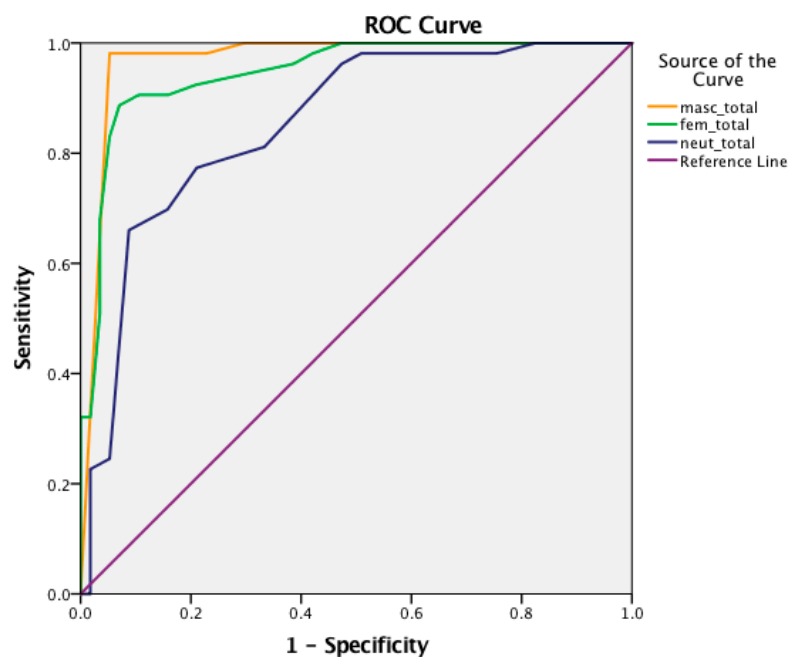


**Figure 1.** Receiver Operating Characteristics (ROC) curve of the GAAGL Assessment in the tasks—monolinguals vs. bilinguals.

**Table 5.** ROC data analysis of the GAAGL Assessment tasks—monolinguals vs. bilinguals.

	Cut-Off	AUC (95% CI)	<i>p</i>
det_real	53.50/54	0.924 (0.868–0.979)	<0.001 *
det_pseud	47.50/56	0.917 (0.866–0.968)	<0.001 *
adj_real	53.50/54	0.828 (0.747–0.908)	<0.001 *
adj_pseud	48.50/56	0.978 (0.958–0.998)	<0.001 *

Notes: det\_real, determiner real words; det\_pseud, determiner pseudo-words; adj\_real, adjective real words; adj\_pseud, adjective pseudo-words; AUC, area under curve; \*  $p < 0.01$ .



**Figure 2.** Receiver Operating Characteristics (ROC) curve of the GAAGL Assessment in the three gender value scores—monolinguals vs. bilinguals.

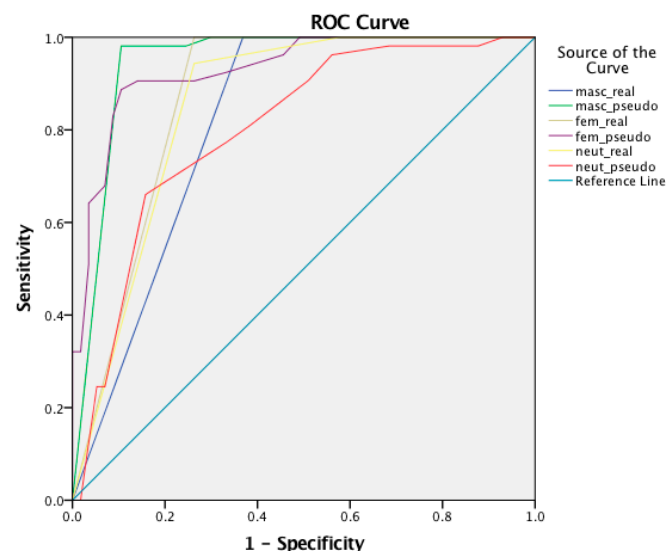
**Table 6.** ROC data analysis of the GAAGL Assessment by gender value score—monolinguals vs. bilinguals.

	Cut-Off	AUC (95% CI)	<i>p</i>
Masc_total	81.50/84	0.969 (0.934–1.00)	<0.001 *
Fem_total	57.50/68	0.947 (0.906–0.987)	<0.001 *
Neut_total	62.50/68	0.852 (0.781–0.924)	<0.001 *

Notes: Masc\_total, masculine total score; Fem\_total, feminine total score; Neut\_total, neutral total score; AUC, area under curve; \*  $p < 0.01$ .

The AUC scores were excellent and ranged from 0.852 to 0.969, with the highest scores attained in the masc\_total scores followed by the fem\_total scores. The cut-off scores in the masculine items were calculated at almost 82, close to the highest end of the measure, while in the feminine nouns they were almost 58/60 and, in the neuter, almost 63/68.

Finally, a separate ROC analysis was run to examine the existence of positive discrimination between the study's groups in the assessment subtypes, as these were defined both by gender and noun type (real or pseudo) ( $3 \times 3$ ). The results showed a high positive discrimination between the two groups in four item subtypes. Figure 3 presents the ROC curve, while Table 7 shows the cut-off scores, the AUC and the  $p$ -value.

**Figure 3.** Receiver Operating Characteristics (ROC) curve of the GAAGL Assessment in the three gender value scores—monolinguals vs. bilinguals.**Table 7.** ROC data analysis of the GAAGL Assessment in the 6 item subtypes scores (by gender and noun type).

	Cut-Off	AUC (95% CI)	<i>p</i>
masc_real	34.50/36	0.816 (0.733–0.899)	<0.001 *
masc_pseudo	45.50/48	0.943 (0.895–0.992)	<0.001 *
fem_real	34.50/36	0.868 (0.796–0.941)	<0.001 *
fem_pseudo	21.50/32	0.930 (0.884–0.977)	<0.001 *
neut_real	34.50/36	0.852 (0.777–0.927)	<0.001 *
neut_pseudo	27.50/32	0.801 (0.718–0.885)	<0.001 *

Notes: masc\_real, real masculine words; masc\_pseudo, masculine pseudo-words; fem\_real, real feminine words; fem\_pseudo, feminine pseudo-words; neut\_real, real neutral words; neut\_pseudo, neutral pseudo-words; AUC, area under curve; \*  $p < 0.01$ .

The AUC scores are excellent, ranging from 0.801 to 0.943, with the highest scores attained in the masc\_pseudo items followed by the fem\_pseudo items. The cut-off scores in

the real items is calculated at 34.5/36 across gender types, while in the pseudo items for the feminine items, it is calculated at 21.5/32, in neuter at 27/5/32, which is higher, while in masculine at 45.8/48.

### 3.3. Cronbach Analysis

Finally, we examined the internal consistency of the GAAGL Assessment by running a Cronbach analysis which estimated the overall internal validity of the test as excellent ( $\alpha = 0.809$ ). Additionally, we run an analysis using the split-half reliability technique which also showed that the test exhibits internal consistency (split-half reliability coefficient = 0.818). Table 8 presents the scores of the item-total correlations and reliability measures of the GAAGL Assessment for its four main tasks.

**Table 8.** Item-total correlations and reliability measures of the four tasks for the gender test.

	Item-Total Correlation
adj_real	0.787
adj_pseud	0.835
det_real	0.750
det_pseud	0.696

Notes: det\_real, determiner real words; det\_pseud, determiner pseudo-words; adj\_real, adjective real words; adj\_pseud, adjective pseudo-words.

As Table 8 shows, the individual reliability measures in the four tasks ranged from 0.696 to 0.835, which reveals that the internal consistency of the test is very good.

## 4. Discussion

This paper aimed at evaluating the psychometric properties of the GAAGL Assessment which comprises four main tasks: two gender assignment tasks where monolingual and bilingual children are asked to provide an agreeing determiner for the real words and pseudo-words and two gender agreement tasks where children provide an agreeing predicate adjective for the real and pseudo-words of the test. Gender tasks tend to reveal differences between monolingual and bilingual children, especially if one or both languages involved grammaticalized gender (Alexiadou et al. 2021; Chondrogianni 2007, 2008; Konta 2013; Kaltsa et al. 2019; Prentza et al. 2019). This is the case for the bilingual participants of the present study, since both L1 Albanian and L2 Greek are grammatical gender languages (Ralli [2005] 2022; Giurgea 2014; Millaku and Topanica-Millaku 2021). In particular, the validity and reliability of the GAAGL Assessment was examined by means of a Cronbach analysis, while for the evaluation of the discriminatory ability of the Assessment, ROC analyses using the Youden index were conducted along with between-group comparisons. The novelty of this paper lies in that it attempts to standardize a grammar-specific language tool with typically developing Greek-speaking monolingual and bilingual populations, something not investigated before. Therefore, it addresses the gap that is detected in relevant research regarding both the evaluation of grammatical measures with typically developing children and the provision of data for the standardization of language tasks with Greek-speaking populations.

Starting with the between-group comparisons across all four tasks, bilingual children scored significantly lower than monolinguals. A closer look of the four tasks of the GAAGL Assessment revealed that the lowest scores for the bilingual children were attained in the adjective\_pseudo task where they had to provide an agreeing adjective predicate for a pseudo-word. When the scores were broken down by gender value, we found that gender attribution, both assignment and agreement, to feminine and neuter nouns is more problematic than gender attribution to masculine nouns. Finally, when scores were broken down by noun type (real or pseudo) and gender value, it was found that the bilingual children score the lowest in the feminine and neuter pseudo-word tasks, both compared to other word types within their group, as well as when compared with



monolingual children. These results agree with data from previous studies, since gender agreement was found to be more challenging for Greek-speaking bilinguals than gender assignment (Chondrogianni 2007, 2008; Kaltsa et al. 2020; Konta 2013, 2019; Prentza et al. 2019), verifying that the syntactic complexity involved in Adj-N structures impacts on gender attribution. Concerning data on real vs. pseudo-words, the results also agree with previous studies on Greek-speaking bilinguals (Kaltsa et al. 2020; Kaltsa et al. 2019; Prentza et al. 2019) supporting the hypothesis that the different mechanisms involved, namely, stored lexical knowledge for the real word (Ralli 2002, [2005] 2022) and exploitation of noun endings for pseudo-words (Tsimpli and Hulk 2013), impact on gender attribution. With respect to gender value scores, we found that performance in masculine nouns is better than performance in feminine or neuter nouns. Given that the masculine nouns of the GAAGL Assessment are associated with noun endings of high predictive value (-is, -as), these results were expected to be in line with previous studies on Greek-speaking bilinguals (see relevant discussion on predictive values Mastropavlou 2006; Varlokosta 2011). However, unlike previous studies (Alexiadou et al. 2021; Kaltsa et al. 2020; Tsimpli and Hulk 2013), neuter nouns in the GAAGL Assessment are not associated with better performance, although they are considered the learner default for (-animate) nouns (see also Anagnostopoulou 2017). The reason for this finding is revealed when scores are broken down both by gender value and noun type; scores on neuter pseudo-nouns are quite low, affecting overall performance in the neuter type. Similar low scores are attained in feminine pseudo-nouns. These noun types share the suffix -i and only orthographic cues can be used to assign gender (-η for feminine and -ι for neuter) posing a challenge for children, as was also shown in previous research (Tsimpli and Hulk 2013).

Next, we turn to the ROC analysis data. Recall that three separate ROC analyses were conducted in the main tasks of the GAGL Assessment and in the Assessment's different sub-scores in order to examine the test's discriminatory ability. Starting with the data from the main tasks analysis, the specificity and sensitivity results are excellent (Chondrogianni and John 2018; Plante and Vance 1994; Rhodes et al. 2019), as are the AUC scores, which ranged from 0.828 to 0.978, with the highest scores attained in the adj\_pseudo task. Regarding the cut-off scores, monolinguals and bilinguals reached different cut-off scores. In the real determiner and real adjective task, they were calculated quite high, at 53.5/54, showing a ceiling effect, while in the pseudo-words, in both the determiner and in the adjective task, they ranged close to 48/56. The cut-off scores can be interpreted as follows: if, for example, an Albanian–Greek bilingual child scores high in the determiner tasks but scores lower, i.e., 46/56 in gender assignment or agreement with pseudo-words, s/he may score lower than monolinguals but still falls within the normal/typical level of bilingual performance, according to the instrument validation data that we present. In the ROC analysis by gender, the results also show high test sensitivity and specificity (Chondrogianni and John 2018; Plante and Vance 1994; Rhodes et al. 2019), while the AUC scores ranged from 0.852 to 0.969, with the highest scores attained in the masc\_total scores. The cut-off scores for the masculine items were calculated close to the highest end of the measure, showing a ceiling effect, while for the feminine nouns they were calculated at nearly 58/60 and, in the neuter, at nearly 63/68.

Finally, in the ROC analysis by gender value and item type, results were also excellent, with the AUC ranging from 0.801 to 0.943, with the highest scores attained in the masc\_pseudo items, followed by the fem\_pseudo items. The cut-off scores in the real items were calculated close to the higher end of the scale across gender types, while in the pseudo items, they were calculated lower (21.5/32 for feminine, 27.5/32 for neuter and 45.8/48 for masculine). These results show that the ability of the test to discriminate between typically developing monolingual and bilingual populations is excellent across tasks and sub-scores. Additionally, gender agreement measures are more sensitive in the detection of group differences than gender assignment measures, which is expected given the more complex syntactic operations involved. Moreover, pseudo-word tasks are also particularly sensitive to group distinctions, given that children cannot use stored lexical knowledge, but they

have to rely on the predictive values of noun endings which are sensitive to language input (Tsimpli and Hulk 2013). Regarding gender value sub-scores, it seems that when blinded to whether the attribution involves a real or a pseudo-word, their discriminatory power shows a slight drop. These results are generally in line with data from between-group comparisons, suggesting that the complexity of syntactic computations in gender agreement processes paired with the challenge of relying only on noun endings for gender attribution in pseudo-nouns are factors that should be exploited in gender assessment in Greek, since they yield valid results with respect to monolingual vs. bilingual distinctions. However, tasks on gender assignment with real words should not be overlooked in gender assessment measures, since the analysis also proved their discriminatory ability, and thus, they could function as a baseline. The ceiling effects attested are expected given that children can rely on to stored lexical knowledge for this task.

The complementarity and the internal consistency of the GAAGLA Assessment, which also indicates that all its main tasks contribute to the tests' validity and should not be eliminated despite the ceiling cut-off scores attained in real word tasks, were shown by the results of the Cronbach's analysis. The Cronbach's alpha analysis evaluated the overall estimated internal validity of the test as excellent ( $\alpha = 0.809$ ) (Nunnally and Bernstein 1994; Ursachi et al. 2015) with the split-half reliability coefficient estimated to be 0.818. Although again the highest scores are arrested in the adjective tasks, especially the pseudo one, the results from the other two tasks are very good, which proves that all four tasks are valid in the examination of gender attribution in Greek.

## 5. Conclusions

This study presented preliminary evidence for the psychometric evaluation and discriminatory ability of the GAAGL Assessment, a gender attribution test designed for Greek-speaking populations. This is the first study to evaluate the GAAGL Assessment with typical school-age children; therefore, it addresses a gap in relevant research regarding the evaluation of measures used with typical Greek-speaking monolingual and bilingual children. The results of the ROC analyses and the relevant cut-off scores showed that the test has excellent discriminatory ability across all its task sub-scores. Additionally, the Cronbach analysis showed a high internal consistency of the assessment across its four main tasks, which shows that tasks targeting gender assignment and agreement with real and pseudo-words are valid in the examination of gender attribution in Greek and, therefore, all four should be included in gender measures. In the future, examination of the GAAGL Assessment with larger samples of Greek-speaking bilingual populations from different L1s (i.e., English, German) could reveal interesting data regarding the psychometric properties of this grammar-specific task. Overall, if there were more validated tests in the Greek language, researchers could profile bilingual children more effectively and properly address their needs.

**Author Contributions:** Conceptualization, A.P., D.T. and I.M.T. methodology, A.P., D.T. and M.K.; software, D.T.; validation, D.T.; formal analysis, D.T. and A.P.; investigation, A.P.; resources, A.P. and M.K.; data curation, A.P., D.T. and M.K.; writing—original draft preparation, A.P., D.T. and M.K.; writing—review and editing, A.P., D.T. and M.K.; visualization, A.P., D.T. and M.K.; supervision, I.M.T.; project administration, I.M.T.; funding acquisition, I.M.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the operational program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) – Research Funding Program: Thales. Investing in knowledge society through the European Social Fund. Thales FP7 Project “Bilingual Acquisition & Bilingual Education: The Development of Linguistic & Cognitive Abilities in Different Types of Bilingualism” (BALED—Award No MIS377313). PI: IM Tsimpli.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of the Pedagogical Institute of Greece (protocol code Φ15/356/42281/Γ1, date of approval: 6 April 2011).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data are part of a larger project (Thales FP7 Project “Bilingual Acquisition and Bilingual Education: The Development of Linguistic and Cognitive Abilities in Different Types of Bilingualism”).

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Note

- <sup>1</sup> Note that the post-analysis of these items did not show any animacy effects in our dataset, hence they are not examined separately in the analysis.

## References

- Agalliu, Fatmir. 2002. *Gramatika e Gjuhës Shqipe I*. Malaka and Tiranë: Botimi i Akademisë së Shkencave.
- Alexiadou, Artemis, Vasiliki Rizou, Nikolaos Tsokanos, and Foteini Karkaletsou. 2021. Gender agreement mismatches in heritage Greek. *Languages* 6: 3. [\[CrossRef\]](#)
- Alloway, Tracy Packiam. 2007. What can phonological and semantic information tell us about the mechanisms of immediate sentence recall. *Memory* 15: 605–15. [\[CrossRef\]](#) [\[PubMed\]](#)
- Anagnostopoulou, Elena. 2017. Gender and defaults. In *A Schrift to Fest Kyle Johnson*. Edited by Nicholas LaCara, Keir Moulton and Anne-Michelle Tessier. Stroudsburg: Linguistics Open Access Publications, pp. 23–31.
- Anastasiadi-Symeonidi, Anna, and Despina Cheila-Markopoulou. 2003. Synchronikes ke diachronikes tasis sto genos tis Elinikis: Mia theoretiki protasi [Synchronic and diachronic trends in the Greek gender: A theoretical proposal]. In *To Genos [Gender]*. Edited by Anna Anastasiadi-Simeonidi, Aggeliki Ralli and Despina Chila-Markopoulou. Athens: Patakis, pp. 13–56.
- Babatsouli, Elena, and Elena Nicoladis. 2019. The acquisition of English possessives by a bilingual child: Do input and usage frequency matter? *Journal of Child Language* 46: 170–83. [\[CrossRef\]](#) [\[PubMed\]](#)
- Balam, Osmer, Usha Lakshmanan, and María del Carmen Parafita Couto. 2021. Gender assignment strategies among simultaneous Spanish/English bilingual children from Miami, Florida. *Studies in Hispanic and Lusophone Linguistics* 14: 241–80. [\[CrossRef\]](#)
- Bianchi, Giulia. 2013. Gender in Italian–German bilinguals: A comparison with German L2 learners of Italian. *Bilingualism: Language and Cognition* 16: 538–57. [\[CrossRef\]](#)
- Bošković, Željko. 2009. Unifying first and last conjunct agreement. *Natural Language & Linguistic Theory* 27: 455–96. [\[CrossRef\]](#)
- Carstens, Vicki. 2010. Implications of grammatical gender for the theory of uninterpretable features. *Exploring Crash-proof Grammars* 3: 31.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge: MIT Press.
- Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A Life in Language*. Cambridge: MIT Press.
- Chondrogianni, Vasiliki, and Nerys John. 2018. Tense and plural formation in Welsh–English bilingual children with and without language impairment. *International Journal of Language & Communication Disorders* 53: 495–514. [\[CrossRef\]](#)
- Chondrogianni, Vicky. 2007. Acquiring clitics and determiners in child L2 Modern Greek. *Selected Papers on Theoretical and Applied Linguistics* 17: 356–66.
- Chondrogianni, Vicky. 2008. Comparing child and adult L2 acquisition of the Greek DP. In *Current Trends in Child Second Language Acquisition: A Generative Perspective*. Edited by Belma Haznedar and Elena Gavruseva. Amsterdam: John Benjamins B.V., pp. 46–97.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Dunn, Lloyd M., and Leota M. Dunn. 1997. *PPVT-III: Peabody Picture Vocabulary Test*. Circle Pines: American Guidance Service.
- Dunn, Lloyd M., Leota M. Dunn, Chris Whetton, and Juliet Burley. 1997. *British Picture Vocabulary Scale 2nd Edition (BPVS-II)*. Windsor: NFER-Nelson.
- Eckes, Thomas. 2017. Setting cut scores on an EFL placement test using the prototype group method: A receiver operating characteristic (ROC) analysis. *Language Testing* 34: 383–411. [\[CrossRef\]](#)
- Fan, Jerome, Suneel Upadhye, and Andrew Worster. 2006. Understanding receiver operating characteristic (ROC) curves. *Canadian Journal of Emergency Medicine* 8: 19–20. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fluss, Ronen, David Faraggi, and Benjamin Reiser. 2005. Estimation of the Youden Index and its associated cutoff point. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 47: 458–72. [\[CrossRef\]](#) [\[PubMed\]](#)
- Giurgea, Ion. 2014. Possible syntactic implementations of the controller vs. target gender distinction: The view from ambigenics. *Language Sciences* 43: 47–61. [\[CrossRef\]](#)
- Hawkins, Roger, and Ianthi-Maria Tsimpli. 2009. *Explaining Target- and Non-Target Performance in the Acquisition of Gender Concord by L2 Speakers*. Utrecht: ISB7.

- Heale, Roberta, and Alison Twycross. 2015. Validity and reliability in quantitative studies. *Evidence-Based Nursing* 18: 66–67. [\[CrossRef\]](#)
- Holton, David, Peter Mackridge, and Irene Philippaki-Warbuton. 1997. *Greek: A Comprehensive Grammar of the Modern Language*. London: Routledge.
- Holton, David, Peter Mackridge, and Irene Philippaki-Warbuton. 2012. *Greek: A Comprehensive Grammar of the Modern Language*. London: Routledge.
- Huang, Francis L., and Timothy R. Konold. 2014. A latent variable investigation of the Phonological Awareness Literacy Screening-Kindergarten assessment: Construct identification and multigroup comparisons between Spanish-speaking English-language learners (ELLs) and non-ELL students. *Language Testing* 31: 205–21. [\[CrossRef\]](#)
- Kaltsa, Maria, Alexandra Prentza, Despina Papadopoulou, and Ianthi-Maria Tsimpli. 2020. Language external and language internal factors in the acquisition of gender: The case of Albanian-Greek and English-Greek bilingual children. *International Journal of Bilingual Education and Bilingualism* 23: 981–1002. [\[CrossRef\]](#)
- Kaltsa, Maria, Ianthi-Maria Tsimpli, and Froso Argyri. 2019. The development of gender assignment and agreement in English-Greek and German-Greek bilingual children. *Linguistic Approaches to Bilingualism* 9: 253–88. [\[CrossRef\]](#)
- Kohnert, Kathryn J., Arturo E. Hernandez, and Elizabeth Bates. 1998. Bilingual performance on the Boston Naming Test: Preliminary norms in Spanish and English. *Brain and language* 65: 422–40. [\[CrossRef\]](#)
- Konta, Eirini. 2013. I kataktisi tis ellinikis apo pedia me mitriki glossa tin tourkiki: Stixia apo tin anomatiki simfonia kai ti morfologia. [The Development of Greek by Turkish-Speaking Children: Evidence from Nominal Agreement and Morphology]. Ph.D. dissertation, Aristotle University of Thessaloniki, Thessaloniki, Greece.
- Konta, Eirini. 2019. The acquisition of gender and the development of derivational morphology in child L2 Greek. *Selected Papers on Theoretical and Applied Linguistics* 23: 226–42.
- Krzanowski, Wojtek J., and Anthony J. Woods. 1984. Statistical aspects of reliability in language testing. *Language Testing* 1: 1–20. [\[CrossRef\]](#)
- Kupisch, Tania, Deniz Akpınar, and Antje Stöhr. 2013. Gender assignment and gender agreement in adult bilinguals and second language learners of French. *Linguistic Approaches to Bilingualism* 3: 150–79. [\[CrossRef\]](#)
- Kurani, Aida, and Anisa Trifoni. 2014. Syntactic Similarities and Differences Between Albanian and English. *European Scientific Journal* 14: 47–62.
- Marinis, Theodoros. 2003. *The Acquisition of the DP in Modern Greek*. Amsterdam: John Benjamins Pub., vol. 31.
- Marinis, Theodoros, Sharon Armon-Lotem, and George Pontikas. 2017. Language impairment in bilingual children: State of the art 2017. *Linguistic Approaches to Bilingualism* 7: 265–76. [\[CrossRef\]](#)
- Martínez-Camblor, Pablo, and Juan Carlos Pardo-Fernández. 2019. The Youden index in the generalized receiver operating characteristic curve context. *The International Journal of Biostatistics* 15: 1–20. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mastropavlou, Maria. 2006. The Role of Phonological Salience and Feature Interpretability in the Grammar of Typically Developing and Language Impaired Children. Ph.D. dissertation, Aristotle University of Thessaloniki, Thessaloniki, Greece.
- Mastropavlou, Maria, and Ianthi-Maria Tsimpli. 2011. The role of suffixes in grammatical gender assignment in Modern Greek: A psycholinguistic study. *Journal of Greek Linguistics* 11: 27–55. [\[CrossRef\]](#)
- Mattheoudakis, Marina, Aspasia Chatzidaki, Christina Maligkoudi, and Eleni Agathopoulou. 2016. Family and school language input: Their role in bilingual children's vocabulary development. *Journal of Applied Linguistics* 31: 49–69.
- Millaku, Shkëlqim, and Xhevahire Topanica-Millaku. 2021. The Albanian and English language gender. *Linguistics and Culture Review* 5: 1197–215. [\[CrossRef\]](#)
- Mitrofanova, Natalia, Olga Urek, Yulia Rodina, and Marit Westergaard. 2022. Sensitivity to microvariation in bilingual acquisition: Morphophonological gender cues in Russian heritage language. *Applied Psycholinguistics* 43: 41–79. [\[CrossRef\]](#)
- Mizuta, Masanobu, Chika Abe, Erina Taguchi, Tsukasa Takeue, Hisanobu Tamaki, and Tomoyuki Haji. 2022. Validation of cepstral acoustic analysis for normal and pathological voice in the Japanese language. *Journal of Voice* 36: 770–76. [\[CrossRef\]](#)
- Mueller Gathercole, Virginia C., Enlli Mon Thomas, and Emma Hughes. 2008. Designing a normed receptive vocabulary test for bilingual populations: A model from Welsh. *International Journal of Bilingual Education and Bilingualism* 11: 678–720. [\[CrossRef\]](#)
- Nunnally, Jum C., and Ira H. Bernstein. 1994. *Psychometric Theory*, 3rd ed. New York: McGrawHill.
- Peña, Elizabeth D., Vera F. Gutiérrez-Clellen, Aquiles Iglesias, Brian A. Goldstein, and Lisa M. Bedore. 2018. *Bilingual English Spanish Assessment (BESA)*. Baltimore: Brookes.
- Perkins, Neil J., and Enrique F. Schisterman. 2005. The Youden Index and the optimal cut-point corrected for measurement error. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 47: 428–41. [\[CrossRef\]](#) [\[PubMed\]](#)
- Pérez-Tattam, Rocío, María José Ezeizabarrena, Hans Stadthagen-González, and Virginia C. Mueller Gathercole. 2019. Gender assignment to Spanish pseudowords by monolingual and Basque-Spanish bilingual children. *Languages* 4: 58. [\[CrossRef\]](#)
- Plante, Elena, and Rebecca Vance. 1994. Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools* 25: 15–24. [\[CrossRef\]](#)
- Prentza, Alexandra, Dionysios Tafiadis, Vasiliki Chondrogianni, and Ianthi-Maria Tsimpli. 2022. Validation of a Greek Sentence Repetition Task with Typically Developing Monolingual and Bilingual Children. *Journal of Psycholinguistic Research* 51: 373–95. [\[CrossRef\]](#) [\[PubMed\]](#)



- Prentza, Alexandra, Maria Kaltsa, Ianthi-Maria Tsimpli, and Despina Papadopoulou. 2019. The acquisition of Greek gender by bilingual children: The effects of lexical knowledge, oral input, literacy and bi/monolingual schooling. *International Journal of Bilingualism* 23: 901–20. [CrossRef]
- Ralli, Angela. 2002. The role of morphology in gender determination: Evidence from Modern Greek. *Linguistics* 40: 519–51. [CrossRef]
- Ralli, Aggeliki. 2022. *Morphology*. Athens: Patakis. First published 2005.
- Raven, John, C. Raven, and John Hugh Court. 1998. *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 2: The Coloured Progressive Matrices*. San Antonio: Harcourt Assessment.
- Revithiadou, A., and V. Spyropoulos. 2013. *A Comparative Study of Albanian-Greek Grammatical Structures. Project 'Education of Immigrant and Repatriate Students' Operational Program of Education and Lifelong Learning (NSRF 2007–2013)*. Thessaloniki: Aristotle University of Thessaloniki Publications, Research Committee. Available online: <http://www.diapolis.auth.gr/index.php/2013-10-17-09-02-51/-52> (accessed on 3 February 2023).
- Rhodes, Stephen, Nelson Cowan, Mario A. Parra, and Robert H. Logie. 2019. Interaction effects on common measures of sensitivity: Choice of measure, type i error, and power. *Behavior Research Methods* 51: 2209–27. [CrossRef]
- Rosselli, Mónica, Alfredo Ardila, M. Gina Navarrete, and Esmeralda Matute. 2010. Performance of Spanish/English bilingual children on a Spanish-language neuropsychological battery: Preliminary normative data. *Archives of Clinical Neuropsychology* 25: 218–35. [CrossRef]
- Stephany, Ursula. 1997. The acquisition of Greek. In *The Cross-Linguistic Study of Language Acquisition*. Edited by Dan Isaac Slobin. Hillsdale: Erlbaum, vol. 4, pp. 183–333.
- Stephany, Ursula, and Anastasia Christofidou. 2008. *The Acquisition of Greek Case, Number, and Gender: A Usage-Based Approach*. Arbeitspapier 55 (Neue Folge) Allgemeine Sprachwissenschaft. Cologne: University of Cologne.
- Theodorou, Elena, Maria Kambanaros, and Kleanthes K. Grohmann. 2017. Sentence repetition as a tool for screening morphosyntactic abilities of bilingual children with SLI. *Frontiers in Psychology* 8: 1–13. [CrossRef]
- Triandaphyllidis, Manolis. 1949. *Concise Modern Greek Grammar*. Thessaloniki: Institute of Modern Greek Studies. Manolis Triandaphyllidis Foundation.
- Tsimpli, Ianthi-Maria. 2003. Clitics and determiners in L2 Greek. Paper presented at 6th Generative Approaches to Second Language Acquisition Conference, Ottawa, ON, Canada, April 26–28; Somerville: Cascadia Proceedings Project, pp. 331–39.
- Tsimpli, Ianthi-Maria. 2014. Early, late or very late?: Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism* 4: 283–313. [CrossRef]
- Tsimpli, Ianthi-Maria, and Aafke Hulk. 2013. Grammatical gender and the notion of default: Insights from language acquisition. *Lingua* 137: 128–44. [CrossRef]
- Ursachi, George, Ioana Alexandra Horodnic, and Adriana Zait. 2015. How reliable are measurement scales? External factors with indirect influence on reliability estimators. *Procedia Economics and Finance* 20: 679–86. [CrossRef]
- Varlokosta, Spyridoula. 2011. The role of morphology in grammatical gender assignment: A psycholinguistic study in Greek. In *Morphology and Its Interfaces*. Edited by Alexandra Galani, George Hicks and Glyn Tsoulas. Amsterdam: John Benjamins, pp. 321–49. [CrossRef]
- Vogindroukas, Ioannis, Athanassios Protopappas, and George Sideris. 2009. *Word-Finding Test*. Chania: Glafki. (In Greek)
- Youden, William J. 1950. Index for rating diagnostic tests. *Cancer* 3: 32–35. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.