Language is a universal human ability, a social tool where specific symbols stand for something else. The development of language moves from simple to complex and is found to persist into adolescence as well. The language development in adolescence is highly sophisticated with advancements in the semantic, syntactic, morphologic and pragmatic aspects of language. Though the linguistic development during this period is subtle, it is considered to be important. The assessment of language in the adolescent period is difficult and highly challenging. The study of language development in adolescence is relatively less, especially in India. With only a handful of standardized language tests available to assess language development in adolescents, identifying language impairment in this population have always been a challenge. Moreover, using the available western adolescent language tests in the Indian scenario has drastic limitations. The current research was taken up to fulfill this need.

The aim of the current research was to develop an adolescent language assessment tool for Indian population between 10 – 16 years, with the objectives of developing an assessment tool to evaluate the adolescent language and administering the developed tool on children between 10 – 16 years.

The research protocol was approved by Institutional Ethical Committee of Kasturba Medical College (Manipal University), Mangalore. This research was carried out in schools within Mangalore, having English as their medium of instruction. This research followed a two stage cluster random sampling research design.

The development of the language assessment tool followed 3 phases. Phase 1 comprised of the designing of the tool along with pilot study, Phase 2 included the administration of the developed language assessment tool, and Phase 3 focused on establishing test validity and reliability measures.

Phase 1 began with identification of the language constructs, tasks, and test items to be used in the tool. Semantic, syntactic and morphologic language constructs were selected consisting of a total of twelve tasks - contrastive relations task (auditory and visual), multiple meanings task (auditory and visual), associated relations task (auditory and visual), convergent naming task (auditory and visual), analogical reasoning task (auditory and visual), morphological derivations task (auditory and visual), double-function words task (auditory), homophones task
(visual), compare/contrast task (auditory), proverbs/idioms task (visual), similes (auditory), and sentence combining (visual). The test items of each of the tasks were prepared based upon the vocabulary and sentences from the prescribed curriculum. Subsequently, tasks were assigned to follow a word or sentence level based presentation under both auditory and visual modalities, except for few tasks which required presentation in specific modality. This was followed by the allocation of the tasks based on the type of response format. Appropriate instructions were framed to suit each task. All tasks followed a standard biserial scoring system, except of the proverbs/idioms task (visual) which followed a categorical scoring system. All the test items of each of the tasks were prepared with caution, with the items being age specific, fitting within the curriculum prescribed, avoiding inter-related and/or inter-locked items, and avoiding the repetition of items and vocabulary within the task and across the modalities. This was followed by subjecting the developed tool for initial validation and revision. The tool was scrutinized by two subject experts in the field of speech language pathology, as well the English teachers of the school on the basis of appropriateness of the task and items. Subsequently, necessary modifications were incorporated and the tool was used for the first pilot study.

The participants for the first pilot study were recruited based on an inclusion and exclusion criteria. An informed consent was obtained from the participants prior to the research conduction. The data collection began with the setting up of the testing environment and following the guidelines for stimuli presentation. The participants in the first pilot study included $N = 90$ individuals between $> 10 \leq 15.11$ years, with 15 participants in each group. The participants of a particular standard received three sets of stimuli – easy, medium, and difficult levels. The obtained data was scored accordingly and items less than 30% scores or more than 80% scores were eliminated from the stimuli set. The ‘sentence combining task (visual)’ was eliminated in view of the varied responses generated by the participants. Following this, the second pilot study commenced with the participants recruited based on a similar inclusion and exclusion criteria that was followed for the first pilot study. The participants consisted of $N = 60$ individuals between $> 10 \leq 15.11$ years, with 10 participants in each group. Similar test room setup and stimuli presentation guidelines were followed which was at par with the first pilot study. The stimuli were presented based on the task level difficulty, and the responses to the test items obtained from the participants were subjected to content validation and reliability related measures. Items receiving an item discrimination power of $>0.3$, item difficulty index of $<40\%$
and >70%, and a Cronbach’s alpha value of >0.70 was considered for inclusion for the final data collection. The tool was subjected to revalidation by subject experts and teachers, and it was ready to be used for the final data collection. Test items under each of the tasks were a total of 5 (except for the proverbs/idioms and similes task) for each group, specific to each modality, totaling to 30 items (5 items X 6 groups) under each task. The proverbs/idioms and similes task which served as a common stimuli for all six groups consisted of 30 items each.

Phase 2 included the administration of the developed language assessment tool. The participants were recruited based on a similar inclusion and exclusion criteria that was followed for the first and second pilot study. The participants for the final data collection comprised of total of \( N = 432 \) participants, divided into six groups. Group I consisted of participants \( (n = 72) \) (36 males and 36 females) within the age range of \( >10 – \leq 10.11 \) years studying in Standard V. Group II consisted of participants \( (n = 72) \) (36 males and 36 females) within the age range of \( >11 – \leq 11.11 \) years studying in Standard VI. Group III consisted of participants \( (n = 72) \) (36 males and 36 females) within the age range of \( >12 – \leq 12.11 \) years studying in Standard VII. Group IV consisted of participants \( (n = 72) \) (36 males and 36 females) within the age range of \( >13 – \leq 13.11 \) years studying in Standard VIII. Group V consisted of participants \( (n = 72) \) (36 males and 36 females) within the age range of \( >14 – \leq 14.11 \) years studying in Standard IX. Group VI consisted of participants \( (n = 72) \) (36 males and 36 females) within the age range of \( >15 – \leq 15.11 \) years studying in Standard X.

The test room setup and stimuli presentation guidelines were at par with the first and second pilot study. The test administration included the presentation of a single stimuli set targeting its age equivalent group. The auditory based tasks were presented first, followed by the visual based tasks. The obtained data was then entered in SPSS version 16 and subjected to statistical analysis. Descriptive statistics was done to analyze the scores that were obtained under each of the tasks under every modality across the groups. A one way ANOVA was done to determine the level of significance \( (p<0.001) \) across the groups for the proverbs/idioms task (visual) and similes task (auditory). Post hoc analysis was done to determine the level of significance \( (p<0.001) \) between the groups.
Phase 3 included establishing the validity and reliability of the developed tool. The process of evaluating the content validity was done during the test construction phase itself. The construct validity of the test was attained by comparing the typically developing adolescents with 60 age matched adolescents with language disorders (10 adolescents with language disorder X 6 groups) using Mann Whitney Test to compare their responses to items and the total scores in every task. Receiver Operating Characteristic analysis was also performed to attain the cut-off scores, sensitivity, specificity, and area under the curve. The process of evaluating the internal consistency was done during the test construction phase itself. In order to test the reliability of the tool, the test was re-administered on 10% of the total sample size (7 participants from each group), after 2 weeks of the initial administration. The coefficient of reliability was found by correlating the scores of the responses of all tasks which were obtained at the two instances using Kappa statistics (agreement between each item in every task) and Intra-class correlation coefficient (agreement between total scores for every task).

The results of each task in both modalities are depicted below.

   - The test items included in the contrastive relations task (auditory and visual) consisted of antonyms which were age specific and increased in complexity with age.
   - The test-retest reliability was high for all groups under both auditory and visual modalities based on Kappa statistics and Intra-class correlation for all the items and the total scores respectively.
   - A good construct validity was attained for all groups under both auditory and visual modalities based on Mann Whitney Test for all the items and the total scores respectively.
   - ROC analysis revealed a moderate-high sensitivity and specificity with cut-off scores within the range of $\geq 1$ - $\geq 3$.

2. Multiple Meanings Task (auditory and visual):
   - The test items included in the multiple meanings task (auditory and visual) consisted of synonyms which were age specific and increased in complexity with age.
• The test-retest reliability was high for all groups under both auditory and visual modalities based on Kappa statistics and Intra-class correlation for all the items and the total scores respectively.

• A good construct validity was attained for all groups under both auditory and visual modalities based on Mann Whitney Test for all the items and the total scores respectively.

• ROC analysis revealed a moderate-high sensitivity and low-high specificity with cut-off scores within the range of \( \geq 1 - \geq 2 \).

3. Associated Relations Task (auditory and visual):

   • The test items included in the associated relations task (auditory and visual) consisted of words which were age specific and increased in complexity with age.

   • The test-retest reliability was high for all groups under both auditory and visual modalities based on Kappa statistics and Intra-class correlation for all the items and the total scores respectively.

   • A good construct validity was attained for all groups under both auditory and visual modalities based on Mann Whitney Test for all the items and the total scores respectively, except for item no. 30V from Group VI which received a poor validity.

   • ROC analysis revealed a moderate-high sensitivity and low-high specificity with cut-off scores within the range of \( \geq 2 - \geq 3 \).

4. Convergent Naming Task (auditory and visual):

   • The test items included in the convergent naming task (auditory and visual) consisted of Aristotelian definitions with the permissible responses being concrete and abstract words, having different word classes increasing with age.

   • The test-retest reliability was high for all groups under both auditory and visual modalities based on Kappa statistics and Intra-class correlation for all the items and the total scores respectively.

   • A good construct validity was attained for all groups under both auditory and visual modalities based on Mann Whitney Test for all the items and the total scores respectively.
• ROC analysis revealed a moderate-high sensitivity and low-high specificity with cut-off scores within the range of ≥1 - ≥2.

5. Analogical Reasoning Task (auditory and visual):
• The test items included in the analogical reasoning task (auditory and visual) age specific which increased in complexity with age.
• The test-retest reliability was high for all groups under both auditory and visual modalities based on Kappa statistics and Intra-class correlation for all the items and the total scores respectively.
• A good construct validity was attained for all groups under both auditory and visual modalities based on Mann Whitney Test for all the items and the total scores respectively; however item no. 13A and 5V from Group I and III respectively received a poor validity.
• ROC analysis revealed a moderate-high sensitivity and low-moderate specificity with cut-off scores within the range of ≥1 - ≥3.

6. Morphological Derivations Task (auditory and visual):
• The test items included in the morphological derivations task (auditory and visual) consisted of bound morphemes which changed with age.
• The test-retest reliability was high for all groups under both auditory and visual modalities based on Kappa statistics and Intra-class correlation for all the items and the total scores respectively.
• A good construct validity was attained for all groups under both auditory and visual modalities based on Mann Whitney Test for all the items and the total scores respectively.
• ROC analysis revealed a low-high sensitivity and moderate specificity with cut-off scores within the range of ≥1 - ≥3.

7. Double-Function Words Task (auditory):
• The test items included in the double-function words task (auditory) consisted of permissible responses which increased in complexity with age.
• The test-retest reliability was high for all groups based on Kappa statistics and Intra-class correlation for all the items and the total scores respectively.

• A good construct validity was attained for all groups based on Mann Whitney Test for all the items and the total scores respectively.

• ROC analysis revealed a high sensitivity and low-moderate specificity with a cut-off score of ≥1.

8. Homophones Task (visual):

• The test items of homophones task (visual) consisted of word choices which were age specific and increased in complexity with age.

• The test-retest reliability was high for all groups based on Kappa statistics and Intra-class correlation for all the items and the total scores respectively.

• A good construct validity was attained for all groups based on Mann Whitney Test for all the items and the total scores respectively.

• ROC analysis revealed a moderate-high sensitivity and low-moderate specificity with a cut-off score of ≥2 - ≥3.

9. Compare/Contrast Task (auditory):

• The test items of the compare/contrast task (auditory) consisted of word choices which were age specific and increased in complexity with age.

• The test-retest reliability was high for all groups based on Kappa statistics and Intra-class correlation for all the items and the total scores respectively.

• A good construct validity was attained for all groups based on Mann Whitney Test for all the items and the total scores respectively, except for item no. 7A from Group II which received a poor validity.

• ROC analysis revealed a moderate-high sensitivity and low-high specificity with a cut-off score of ≥2 - ≥3.

10. Proverbs/Idioms Task (visual):

• One-way ANOVA revealed a main significant effect across the groups indicating an overall development in the comprehension of proverbs and idioms.
Post hoc analysis revealed a significant difference only between Group II and III, and Group IV and V, which indicated that not all groups exhibited a significant difference between them.

The test-retest reliability was high for all groups based on Kappa statistics and Intra-class correlation for all the items and the total scores respectively.

A good validity was attained for the total scores of the task using Mann Whitney Test, except for Group II which attained a poor validity.

ROC analysis revealed a moderate-high sensitivity and low-moderate specificity with a cut-off score of ≥20 - ≥38.

11. Similes Task (auditory):

One-way ANOVA revealed a main significant effect across the groups indicating an overall development in the comprehension of similes.

Post-hoc analysis revealed no significant differences between any of the groups, indicating that an improvement in the interpretation of these expressions was not evident when compared between groups.

The test-retest reliability was high for all groups based on Kappa statistics and Intra-class correlation for all the items and the total scores respectively.

Only certain test items received a good validity based on the Mann Whitney Test. However, a good validity was attained for the total scores of the task using Mann Whitney Test.

ROC analysis revealed a moderate-high sensitivity and low-moderate specificity with a cut-off score of ≥2 - ≥14.

Clinical Implications

The developed language assessment tool is a norm-referenced and a criterion referenced test which can be used in clinical settings to gain insight about task specific problems encountered by adolescents with language disorder. Consequently, by knowing the exact nature of the adolescent’s language disorder, a highly structured and tailor-made intervention plan can be prepared to help them communicate effortlessly.
Limitations and Future directions

The developed language tool was limited to assess the semantic and morphological aspects of language. The tool which is developed using the state board syllabus was aimed to assess adolescents between 10 – 16 years in Mangalore. The present research included adolescents with a general language disorder. It would be interesting for the future research to explore the syntactic and pragmatic aspects of language to provide an overall developmental pattern of the adolescent language. There is scope for future research to extend the adolescent age range and develop regional based norms. Future studies are also directed to check the pattern of language development in adolescents with specific language disorders.