DEVELOPMENT AND VALIDATION OF HINDI AUDITORY NAMING TEST (HANT)

Chhavi Tyagi

Reg. No: P01II24S123010

A Dissertation Submitted in Part of Fulfilment of the

Degree of Master of Science

(Speech- Language Pathology)

University of Mysore

ALL INDIA INSTITUTE OF SPEECH AND HEARING,

MANASAGANGOTHRI, MYSURU -570006

JUNE, 2026

CERTIFICATE

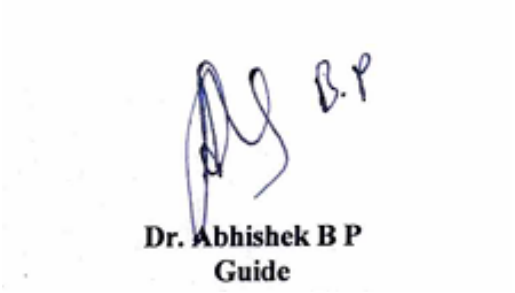
This is to certify that this dissertation entitled **"Development and Validation of Hindi Auditory Naming Test (HANT)"** is a bonafide work submitted as part of fulfilment of the degree of Master of Science (Speech-Language Pathology) of the student with Registration Number P01II24S123010. This has been carried out under the guidance of a faculty of this institute and has not been submitted earlier to any other university for the award of any other Diploma or Degree.

Mysuru
June, 2026

Dr. M. Pushpavathi
Director
All India Institute of Speech and Hearing
Manasagangothri, Mysuru-570 006

CERTIFICATE

This is to certify that this dissertation entitled **"Development and Validation of Hindi Auditory Naming Test (HANT)"** has been prepared under my supervision and guidance. It is also being certified that this dissertation has not been submitted earlier to any other university for the award of any other Diploma or Degree.



Mysuru
June, 2026

Dr. Abhishek B P
Guide

Assistant professor of Language Pathology, Centre of Speech-Language Sciences
All India Institute of Speech and Hearing
Manasagangothri
Mysuru-570 006

DECLARATION

This is to certify that this dissertation entitled **" Development and Validation of Hindi Auditory Naming Test (HANT)"** is the result of my own study under the guidance of Dr. Abhishek B P, Assistant Professor of Language Pathology, All India Institute of Speech and Hearing, Mysuru and has not been submitted earlier to any other university for the award of any other Diploma or Degree.

Mysuru
June, 2026

Registration no. P01II24S123010

Acknowledgements

Hare Ram, Hare Krishna, First and foremost, I am grateful to Bhagwanji
And to the Father who taught me to walk.

I wish to dedicate this dissertation to the most hardworking, ambitious and dedicated human beings (infact, superbeings), my Mumma (***Smt.Balbir Kaur****) and Papa* ***(Shri Ramavtar Tyagi****). Your sacrifices, unconditional love, and unimaginable hardwork have brought me here.I will always remain in debt to my parents.*

I express my heartfelt gratitude to my guide and a strong pillar of strength, ***Dr. Abhishek B P****, for constant encouragement and unwavering support that sir has provided me in this journey. Sir, you stood as a pillar whenever the path seemed uncertain, Under your guidance I have developed interest and confidence in doing research at the same time I have gained valuable life lessons which include consistency, punctuality and hardwork.*

I sincerely thank ***Dr. M. Pushpavathi*** *mam (Director of AIISH), for providing me this opportunity to strengthen my research and clinical skills, during my time in your esteemed Institute.*
I am grateful to ***Kapila*** *for providing a safe, comfortable and memorable space to live for 2 years, and giving me a home away from home.*

I would like to express special thanks to my fiancé, ***Mr. Jitendra Kumar Saini****, for constant support and encouraging words that provide strength in times of self-doubt. Your presence has given me courage whenever I needed it the most.*

I thank my brother, ***Arun Tyagi****, for always staying present for parents and for me whenever the situation demanded strength and support*

I thank my full of life kind of girl, ***Manisha****, for making life less complicated here and giving me a company in making the days here more memorable, your friendship is something I will always cherish in life.*

*On the similar terms, I am grateful to have found **Manasi** as such a great friend who have not only supported me as a friend but also inspired me and even guided me at some points.*

*I am also grateful to **Tanya and Vishnu,** for giving me an ear when I needed, friendship like yours is hard to find. You guys have provided me moral support to reach here and to keep moving. You truly deserve a big thanks for dealing with my behaviour as well.*

*I also acknowledge **Samritha** as a compassionate ad supportive dissertation partner who made this journey smoother with her cooperation and kindness.*

*I am sincerely thankful to **Jahnavi mam**, for always helping me out when I needed, specially in technical issues, with soo much patience.*

*Thanks to my lovely **classmates**, you have directly or indirectly taught me many lessons and helped me develop at an individual level as well.*

*I am thankful to all the **professionals** who contributed to the validation and reliability testing process, as well as the participants for giving your precious time for this study.*

*Last but not the least I and grateful for **technology and digital tools**, which made this process more efficient and manageable.*

Without the support, love, and guidance that I received, this journey would not become possible.

Thank you so much!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

List of Frequently Used Abbreviations

Abbreviation	Full Form
AABR	Automated Auditory Brainstem Response
ABR	Auditory Brainstem Response
ANT	Auditory Naming Test
AQ	Aphasia Quotient
BDAE	Boston Diagnostic Aphasia Examination
BNT	Boston Naming Test
EEG	Electroencephalography
ERP	Event-Related Potential
fMRI	Functional Magnetic Resonance Imaging
HANT	Hindi Auditory Naming Test
I-CVI	Item-wise Content Validity Index
LEAP-Q	Language Experience and Proficiency Questionnaire
MEG	Magnetoencephalography
MoCA-H	Montreal Cognitive Assessment – Hindi Version
PET	Positron Emission Tomography
RSN	Rapid Serial Naming
SLP	Speech-Language Pathologist
WAB	Western Aphasia Battery
WEAVER++	Word-form Encoding by Activation and VERification Model
ms	Milliseconds

ABSTRACT

The process of retrieving the lexical information is one of the essential component of spoken language production and it plays an important role in everyday communication. Traditional naming tasks may not adequately reflect the real life communication demands where language is mainly processed through auditory mode. Auditory naming tasks require additional cognitive- linguistic processing Even though there is an existing clinical importance of auditory naming, there is no standardized auditory naming tool available in Indian languages, particularly in Hindi.

Aim and Objectives: *The present study aimed to develop and validate a Hindi Auditory Naming Test (HANT) for assessing lexical retrieval abilities in neurotypical Hindi- speaking individuals. 1) To develop and validate the Auditory Naming Test, 2) To conduct field testing across three age groups (18–30, 31–45, and 46–60 years) and examine performance variations among them, 3) To establish inter-rater and test–retest reliability using 10% of the collected data.*

Method: *The study was conducted in three different phases. In the first phase, 31 auditory stimulus questions were developed in hindi using linguistically and culturally appropriate target words which tended to represent varying lexial frequencies and semantic categories. In the second phase, content validation was carried out by five experts in Speech- Language Pathology. In the third phase, the finalized stimuli were administered on 60 neurotypical Hindi-speaking participants divided in 3 age groups. Test- retest and Inter-rater reliability measures were also carried out on 10% of the collected data.*

Results: *The developed HANT demonstrated satisfactory content validation for 30 items, whereas 1 item received negative feedback from the validators and hence*

it was removed. Participants across all the age groups were able to comprehend and respond appropriately to the auditory naming stimuli in Hindi. High-frequency and concrete words showed greater naming accuracy compared to low-frequency and abstract items. Younger individuals performed better compared to older age group individuals. Qualitative analysis showed variations in response patterns efficiency across different complexities and types of stimuli. Reliability analysis demonstrated excellent inter-rater agreement and test-retest reliability.

Discussion: *The findings of the present study support the clinical relevance of using auditory naming tasks that are involved in HANT to assess lexical retrieval in Hindi-speaking adults. Auditory naming puts greater demands on aspects like auditory comprehension, semantic processing, and working memory, which are generally not assessed through visual confrontation naming. Hence, this test provides a more ecologically valid measure of everyday communication needs. Lexical frequency and semantic complexity influence the naming performance, which aligns with existing models of lexical access. The inclusion of culturally relevant and linguistically appropriate stimuli further enhanced the ecological validity of the test for the Indian population.*

Conclusion: *This study highlighted the importance of developing culturally relevant and linguistically appropriate assessment tools for the Indian population in order to improve the ecological validity and clinical applicability of language assessment procedures. It also highlights the age related trends in naming accuracy.*

Chapter 1

Introduction

Naming objects is one of the most fundamental linguistic abilities in humans and forms a core component of spoken language production. In everyday communication, individuals continuously retrieve words to label objects, actions, and concepts in order to convey meaning. From a psycholinguistic point of view, naming is not a single-step process. Infact, it involves a complex multistage cognitive-linguistic system.

Levelt (1989) defined naming as the process of retrieving a lexical item from the mental lexicon to express a concept during speech production. The mental lexicon refers to the internal storage of words in the brain that contains semantic, grammatical, and phonological information about each lexical item. For an instance, when a person encounters a dog, the concept of the animal is first activated, which is then followed by the retrieval of the word dog, and finally, its phonological encoding and articulation occur.

According to Levelt's model of speech production, successful naming requires coordinated functioning across multiple processing stages, that is; conceptual preparation, lemma retrieval, and phonological encoding (Levelt, Roelofs, & Meyer, 1999). Conceptual preparation involves activating the intended meaning. Lemma retrieval involves selecting the correct lexical entry along with its grammatical properties, and phonological encoding involves retrieving and sequencing the sounds for articulation. These stages are important to be noted since any breakdown at these stages can lead to naming errors, and can be informative during clinical assessment.

Goodglass and Wingfield (1997) in their study described naming as an ability to acesss and produce spoken word in response to a picture or verbally described item. Visual and auditory, these are the two common modalities through which naming is assessed using different types of stimuli.

Confrontation naming is one of the widely used approaches, in which individuals name visually presented objects or pictures. It involves, what we call as, our primary sense i.e vision, to recognize an object and also for semantic activation (Goodglass, Kaplan, & Barresi, 2001). On the other hand, responsive naming involves providing a spoken verbal stimulus, and expecting a single word response from the individual. For example, answering the question "What do you use to cut vegetables?" requires understanding of the sentence by the listener and retrieve the word knife.

Responsive naming imposes additional cognitive demands when we look from a psycholinguistic point of view. To break it down, first the listener must decode the auditory signal, analyse sentence structure, extract semantic intent and maintain the information in working memory for retrieving information from the lexicon (Gainotti, 2014). Only after this process, a lexical item is selected and produced appropriately. Thus, we can say responsive naming is particularly sensitive to identify any disturbances in auditory comprehension, working memory and lexical retrieval stages of this process.

Understanding of this phenomenon is important for clinical scenarios, particularly in aphasias. Aphasia involves impairment in multi-stage language processing, where deficits may not equally affect each particular stage (Nickels, 2002). Comprehensive aphasia batteries such as the Western Aphasia Battery also include a separate responsive naming subtest that uses functionally relevant verbal

questions to elicit a response (Kertesz, 2006). Hence, responsive naming task is said to closely mimic real life communication demands, where mostly the language is processed in an auditory verbal context and not visual.

Recent research has suggested that responsive naming shows higher ecological validity, which means its more practical in nature and reflects real-life communication demands (Fergadiotis & Wright, 2011; Pichora-Fuller & Singh, 2006). In daily conversational tasks, individuals hear a verbal message, which they interpret after auditory reception, and respond verbally in real time. Similarly, auditory tasks may uncover deficits that remain unidentified through picture-based assessments, particularly in individuals with auditory comprehension or working memory issues (Thompson & Hier, 1998).

In addition to the mode of message transfer, performance in naming tasks is also influenced by the type of lexical item itself, that means if the words or lexical item represent a concrete or abstract concept. Concrete words refer to tangible entities that can be perceived through the senses (e.g., apple, chair, dog), whereas abstract words represent intangible concepts (e.g., justice, freedom, honesty). The Dual Coding Theory (Paivio, 1991) proposes that concrete words benefit from dual representation both verbal and imaginal, whereas abstract words rely mainly on linguistic representation. This results in the well-documented concreteness effect, where concrete words are named faster and more accurately than abstract words (Binder et al., 2005).

Another important semantic distinction in naming research involves living versus non-living categories. Neuropsychological studies have reported category-specific naming impairments, suggesting that semantic knowledge may be organised

in partially distinct neural networks. Several studies have shown that individuals with brain damage may demonstrate greater difficulty naming living things (e.g., animals, fruits) compared to non-living objects (e.g., tools, vehicles) (Warrington & Shallice, 1984; Gainotti, 2000). This pattern has been explained by the sensory–functional theory (Warrington & Shallice, 1984) , which proposes that knowledge of living things relies more heavily on sensory and perceptual features, whereas non-living objects depend more on functional and action-related information. Because functional information tends to be more robustly represented, non-living vocabulary is often retrieved more efficiently and preserved better following brain damage. Incorporating such semantic distinctions is therefore essential when developing naming assessments.

1.2 Need for the Study

Traditional Visual Naming Tests primarily assess lexical retrieval using picture-based stimuli. However, these tests may not adequately capture the word-finding difficulties experienced in everyday conversation, where communication is predominantly auditory. Visual stimuli alone may therefore not provide a complete picture of lexical retrieval abilities in a person.

An auditory naming test can reflect naturalistic real life communication demands in a better way. This approach can help in differentiating modality specific impairment in language abilities which may also guide clinical decision making. It is particularly useful for identifying individuals who are able to perform better in visual naming but experience difficulty in lexical retrieval when spoken input is given.

Unlike traditional tests that rely solely on accuracy, auditory naming can also incorporate measures such as reaction time, offering greater diagnostic sensitivity. Developing and validating an ANT has the potential to improve the accuracy of word-retrieval assessment, enhance diagnostic precision, and support more targeted intervention planning for individuals with language processing difficulties.

1.3 Aim:

The present study aims to develop and validate an Auditory Naming Test in Hindi to assess lexical retrieval in the neurotypical population across age groups using auditory modality.

1.4 Objectives:

1. To design the content of the Auditory Naming Test and validate it to ensure it is linguistically appropriate and culturally relevant.
2. To conduct field testing on a controlled sample to examine performance variations across three age groups (18-30, 31-45, and 46-60) and to assess the performance variations across the three age groups.
3. To determine inter-rater and test–retest reliability by analysing 10% of the collected data.

Null Hypothesis (H_0) for the objectives two and three is as follows:

- There is no statistically significant difference in the performance of participants of the 3 age groups.
- There is no significant inter-rater reliability and test–retest reliability for the Auditory Naming Test in Hindi.

Chapter 2

Review of Literature

1.1 Lexical Access

The term, Lexical access, refers to a process by which speakers retrieve words from their mental lexicon during the process of speech production. The mental lexicon is a long-term memory store containing representations of a speaker's vocabulary, including information about word meaning, syntactic properties, and phonological forms (Levelt, 1989). If the listener is able to access this lexical information efficiently, he/she may demonstrate fluent language use, this allows the person to select contextually appropriate words rapidaly with minimum conscious effort. In more naturalistic conversation, this process is more automated and highly efficient, which takes place in very short span of time. Typically under 200 ms. (Levelt, 1999).

The process of lexical access is generally understood to occur in distinct but closely connected stages. According to Levelt, 1999 lexical access begins with conceptual preparation, in which the speaker activates relevant semantic features of the intended message. This conceptual activation leads to selecting a corresponding lemma, an abstract, modality-independent representation of a word that contains syntactic and semantic features but lacks phonological form. Once the appropriate lemma is selected, the system retrieves the phonological form or lexeme, comprising the actual sounds needed to articulate the word. This staged architecture of lexical access, conceptual activation, lemma node activation, and phonological encoding, forms the basis of several psycholinguistic models.

Naming is a key linguistic task that taps the various processes involved in the lexical access from retrieving the correct word for a given concept or stimulus, such as a picture or object, to producing the 'specific label'. During naming, the conceptual representation of the target activates multiple parallel lemmas that share semantic features. Among these, the lemma that surpasses a critical activation threshold is selected for further processing (Levelt, 1999). After the target lemma is selected the phonological form for that lemma is retrieved, and it is prepared for articulation. Naming, in fact, illustrates almost the whole lexical access process, which mirrors the way in which meaning is converted into real-world output.

There are several models of lexical access which address differences in activation and the relationship between linguistic levels and naming. The Discrete Serial Model (Levelt, 1999) is the most recent model of spoken. A word production system that is feed-forward (or a step-by-step) system in which one stage of production proceeds before another. This model suggests that speech begins with conceptual preparation the act of the speaker preparing an intention and choosing the desired meaning of what he or she wants to say. It is then a preverbal message that triggers the second stage of lemma selection, in which the mental lexicon is consulted to find the correct lexical item, based on meaning and grammatical constraints but not its sound form. When the lemma is fully accessed, the next stage is phonological encoding, in which the phonological form of the word is retrieved including its phonemes, syllable structure, stress pattern and segmental details. Then the system continues to syllabify, that is, to organise phonological units into pronounceable syllables, before encoding the phonetic plans, that is, the articulatory plans for the next stage of the system, namely, phonetic encoding. These are composed using motor programs that are stored.The last stage is articulation, in which the commands to the

motor are sent to the output, the spoken output. One of the major points of this model is that all of these steps are independent of each other, there is little overlapping of steps, and there is no feedback from later steps to earlier steps which would cause a circular progression from meaning to sound.

This strict seriality allows us to understand the linguistic information that is structured in the production of speech with ease and gives us different points on which the breakdowns can take place in people with language and speech impairments. Evidence for this model comes from a study done by Damian and Martin (2001), who used modified picture–word tasks to determine where semantic interference occurs in the speech production process. They used a blocked-cyclic naming task, in which sets of semantically related pictures and sets of unrelated pictures were presented in distinct blocks, with participants naming sets of semantically related pictures and sets of unrelated pictures repeatedly, with the manipulation of the modality of distractors (auditory vs. visual), the timing of the stimulus onsets.

The study found that semantic interference reliably occurred only when lexical access was required, and not when pictures were named using non-lexical cues (e.g., naming digits or artificial objects with pre-learned labels). It was found that, interference was strongest in early cycles and diminished as lexical competition reduced over repetitions. These patterns indicated that the interference develops during lexical selection, not at the conceptual or phonological levels. This complements the Discrete Serial Model by showing that lexical selection is a distinct processing stage, and disruptions at this level do not bleed into the phonological stage.

In contrast to the Discrete Serial Model, the Spreading Activation Model proposed by Dell in 1986 describes lexical access as an interactive process in which

activation flows bidirectionally between the semantic, lexical and phonological levels. The spreading activation model framework explains that multiple lexical and phonological representation levels may be partially activated simultaneously, and the feedback between these levels allows the dynamic adjustment during the word selection process. This model also explains the patterns observed in speech errors, such as phonological blends or semantic substitutions, by comparing them to competing activations across the levels of representation.

Support for this model is provided by Indefrey and Levelt (2004), who carried out a meta-analysis of chronometric and neuroimaging research on word production. They compared Data from 82 Experiments, including picture naming, word generation, and brain imaging (fMRI and EEG). According to their analysis, conceptual preparation takes place at about 175 milliseconds after stimulus onset, the selection of the lemma at about 200 milliseconds, and the encoding of the phonology at about 275 milliseconds. The different activation times suggest sequential lexical access with well-defined temporal boundaries between stages. The Discrete Serial Model and the Spreading Activation Model share similar goals but do not describe the organization and interaction of processing stages in the same fashion.

There is a notable similarity in that both frameworks give a sense that there are many levels involved in naming that include conceptual, lexical, and phonological levels, and that word production involves the coordinated activation of these levels. Both also appeal to the evidence of behaviour (semantic interference and speech-error patterns) in their argumentation.

The key difference is in the interaction between these stages. According to the Discrete Serial Model, the process is strictly feed-forward and step-by-step, in which the work of one level must be finished before the next level is started and there is no feedback from phonology to. semantics or lemmas. The opposite view, the Spreading Activation Model proposes a parallel and interactive activation process in which there can be more than one activation at a given time and information can flow both forward and backward between levels. This is an interactive account that addresses the reasons for the influence of phonological or semantic competitors on word choice and for the occurrence of mixed error types. As such, the model by Levelt presents clearly defined stages and independence between them, while the model by Dell accentuates continuous interaction and mutual influences. Later models try to reconcile these two theoretical contrasts, that of serial and interactive approaches to lexical access.

The Spreading Activation Model proposed by Dell (1986) proposes that the production of spoken words is an interactive system where activation is continually distributed across various levels of the language, such as semantic features, lexical items and phonological segments. According to this model, a speaker might want to convey an idea to an audience, and the activation of the semantic nodes would increase until it reached the lexical items, and from there it would go to the phonological forms of the items, which would establish a dynamic pattern of activation in the network. Importantly, the model proposes that this activation is bilateral, that is, that the activation of phonological information could influence the activation of the lexical information and vice versa, that the activation of lexical information could strengthen or weaken the activation of the semantic information. Due to this bi-directional flow, processing may occur at multiple levels simultaneously and multiple candidates may be partially activated at each level, not only the intended candidate. “dog”).The most

highly activated phonological sequence is then selected at retrieval time, and speech output is produced, giving a model that can explain common speech errors like semantic substitutions (e.g., “cat for “dog”). Errors include phonological errors (“dog” for “cat”) and mixed errors (“mat” for “cat” that are both semantically and phonologically related.

The Spreading Activation Model does not assume a strict sequence of different stages of speech production but instead focuses on a very interactive and flexible architecture: the more an element is activated, the more likely it is to be selected; feedback loops are active in the production of the final spoken output. This interactive structure describes the variability, error patterns and the ease of production that we observe in real speech production.

The spreading activation model has been supported by an examination of speech production errors, as in Dell et al. (1997). The study comprised of adult speakers of English from a variety of natural conversational and controlled laboratory based naming tasks. In the laboratory, the participants (about 20 adults) were asked to name the pictures, fill in the word lists or describe the pictures in a scene while their speech was recorded. The spontaneous speech samples in the conversational data were transcribed and examined for spontaneous phonological and semantic errors.

Errors were categorized based on their level of language: semantic (e.g., referring to a cat instead of a dog), phonological (e.g., “dot” for “dog”) and mixed/blended (e.g., “doy” for “dog” and “boy”). The researchers subsequently analyzed the types of errors and their distribution across contexts statistically.

They discovered that the errors were not random, but that the same error types were found in spontaneous and elicited speech. This revealed that both the semantic

and phonological processes were simultaneously and mutually activated during the production of the words, and this has supported the theory of the interactive spreading activation lexical access model proposed by Dell. Additional support for this was given by Rapp and Goldrick (2000) who examined three aphasic speakers on the basis of their word production tasks. They examined the relationship between semantic and phonological processes in people with aphasia in order to gain insight into lexical retrieval function in a system whose components are limited. They used participants with known word retrieval problems for word production tasks such as picture naming, word repetition and oral reading.

The researchers thoroughly examined the speech errors produced by participants in terms of their type and their pattern. Errors were classified as semantic (e.g., saying “cat” for “dog”) or phonological (e.g., saying “dot” for “dog”). They also looked at mixed errors answers in which there was both a semantic and phonological resemblance (e.g., “rat” for “cat”). They found that semantic and phonological errors frequently occurred together and not separately when they compared the frequency and type of such errors.

This suggested that activation at the semantic level may persist and impact on phonological encoding after the point at which lemma selection has started. The findings were consistent with the spreading activation model of lexical retrieval, with their results providing good evidence for the two-way flow of information from semantic to phonological and vice versa. This model argues against a strictly serial view of word production, and highlights how multiple processing levels interact and provide feedback throughout the production of words.

Caramazza and Hillis (1990) suggest a model of lexical access, called the Semantic Feature Model, that focuses on the detailed semantic representations, consisting of distinct and shared conceptual features. Here, every concept is not simply a single abstract node, but a network of semantic attributes, such as [+living] or [+animal] or [+has fur] or [+barks] or [+pet] for the concept of "dog". If the speaker plans to refer to an object, activation of the lexical features extends over these properties, and lexical selection is influenced by the degree of match between the feature set activated by the speaker and the feature set of the candidates. Co-activated words compete, particularly when their activation is weak, degraded, or spread over a number of similar words; and if activation of a word is weak, degraded, or spread over a number of similar words, its co-activation enhances competition, making selection more difficult and, when semantic information is not complete, or is overlapping or noisy, words become more difficult to name. This model therefore considers fine-grained semantic effects of naming errors, which are errors made within closely-related categories, category-specific impairments, and the effect of feature typicality on retrieval. The Semantic Feature Model was found to have a number of important characteristics in its ability to explain lexical access: it highlights the organisation and distinctiveness of semantic features that is relevant to lexical access, and it is especially valuable in explaining how small conceptual distinctions are used to ensure accurate word selection during speech production.

In support of this, Caramazza and Shelton (1998) examined individuals with category-specific naming deficits through single-case studies involving patients with focal brain lesions. Participants performed picture-naming tasks in which they were shown images from different semantic categories, such as animals, fruits, tools, and

vehicles, and were asked to name them aloud. They also completed word-to-definition matching tasks, where they matched verbal definitions to corresponding words.

Accuracy and error types were analyzed across categories, and control tasks were included to rule out perceptual or phonological problems. The researchers found that some participants could name non-living objects like tools and vehicles accurately but had difficulty naming living things such as animals and fruits. This pattern suggested that the semantic system is organized based on distributed feature networks, where different conceptual categories rely on distinct sets of semantic features, i.e semantic system (our brain's meaning network) is organized based on different kinds of features, like sensory vs. functional, rather than being stored as one big group of words.

Both the Spreading Activation Model and the Semantic Feature Model highlight that lexical access is influenced by patterns of activation within the semantic system, and both argue that multiple lexical candidates become active during naming. They also share the view that competition between related items plays an important role in determining the final word produced. However, they differ in how they explain the source and organisation of this activation. The Spreading Activation Model emphasises a fully interactive framework in which activation flows continuously and bidirectionally between semantic, lexical, and phonological levels, and it accounts for naming behaviour largely through the strength and timing of activation across these interconnected layers. In contrast, the Semantic Feature Model focuses more narrowly on the internal structure of semantic representations themselves, proposing that words are selected based on how well their feature sets match the conceptual pattern activated at a given moment. While Dell's model explains errors through interactions

across levels, Caramazza and Hillis (1990) argue that naming difficulties often arise from fine-grained feature overlap within specific conceptual categories. Thus, the two models complement each other: one describes how information flows through the system, while the other explains how semantic content is structured, together offering different perspectives on why semantically related words compete during lexical retrieval.

WEAVER++ model (Roelofs, 1992) combines behavioural and chronometric data to account for the temporal dynamics of word production. It has the same feedforward structure as Levelt's model, but it adds a very constrained activation mechanism, in which lexical selection is controlled by very specific timing constraints. The model predicts response latencies in naming tasks, particularly in experimental distractor manipulations. Semantic distractors, for example, slowed down response times when they were presented immediately prior to the naming target, whereas phonological distractors can aid retrieval when presented at the right moment (Bürki & Madec, 2022).

These results validate the importance of the timing of activation at each level and validate models that incorporate detailed temporal architectures. Roelofs (1992) continued with picture-word interference experiments, in which a target picture (e.g., a cat) was presented together with distractor words that were semantically related to the target (e.g., dog) or unrelated (e.g., table). were either semantically or phonologically related. The results indicated that the semantically related distractor (e.g., “dog”) slowed the naming of the picture, an effect referred to as semantic interference. When the distractor shared, however, Naming was faster, demonstrating phonological facilitation, for phonological overlap (e.g., “cap”). More recent studies

by Bürki and Madec (2022) reproduced these findings with reaction-time analysis and EEG, showing that the timing between the presentation of the stimulus and the activation of each stage is crucial for the retrieval efficiency.

While earlier models suggested a sequential process of lexical access, more recent evidence suggests a partially parallel process. In their study, Strijkers and Costa (2011) used electroencephalography (EEG) to investigate the time course of lexical access during speech production. Participants were asked to complete a picture-naming task, which involved showing them pictures on a screen and naming them as fast and accurately as they could. The researchers used event-related potentials (ERPs) to see when various linguistic processes took place in the brain. In particular, they concentrated on neural elements involved in lexical-semantic retrieval and phonological encoding. The EEG data showed that brain regions associated with phonological processing (e.g., left superior temporal gyrus) were activated very early, prior to the completion of lemma selection. This result contradicted the serial model and favored a cascaded model of lexical access, in which the semantic and phonological stages are not completely distinct but rather overlap in time.

Indefrey and Levelt (2004) carried out a meta-analysis of numerous neuroimaging (PET and fMRI) and electrophysiological (EEG and MEG) studies related to speech production. Their aim was to estimate the temporal and spatial sequence of cognitive processes involved in naming. By collecting the results from various studies using the picture-naming paradigm, they proposed an approximate timeline for each processing stage conceptual preparation (0–200 ms), lemma retrieval (200–275 ms), phonological encoding (275–450 ms), and articulation (>450 ms). However, the overlap observed between lemma retrieval and phonological encoding

stages suggested that these processes might operate in parallel instead of strictly one after another. This reinforced the idea of a cascaded flow of activation during lexical access.

Lexical access models provides a basic framework for explaining various linguistic processes, especially naming. Naming any picture involves a sequence beginning with conceptual activation, lemma node activation, and phoneme activation. In the stage of conceptual activation, relevant semantic features are activated. During lemma activation, these features match conceptual representations, activating multiple lexical nodes that share these conceptual features. Hence, several lexical nodes get activated simultaneously.

Among these, the node corresponding to the target picture item is selected based on its activation threshold; the lexical node that attains a higher threshold will be selected. Following this, the corresponding phonemes of the selected target item are activated. This is how lexical access models connect to the process of naming.

During a naming task, such as when a person names a pictured object, lexical access is initiated through cascading stages, including conceptual activation, lemma selection, and phonological or l\ent valphoneme activation.

Conceptual activation marks the initial stage of lexical retrieval, wherein the visual or auditory stimulus (e.g., a picture or description) activates a set of semantic features associated with the target concept. For instance, viewing an image of a "dog" might activate features such as "animal," "four-legged," and "barks" (Caramazza & Shelton, 1998). This activation is distributed and can spread to related conceptual nodes that share overlapping semantic attributes, leading to the simultaneous activation of multiple related lexical items (Dell, 1986; Levelt, 1999).

Following this, lemma activation occurs. A lemma is an abstract, pre-phonological representation of a word that contains syntactic and semantic information but lacks phonological detail (Kempen, 1983; Levelt, 1999). From the pool of activated lexical items, the lemma that most closely matches the intended concept is selected for further processing. The selection mechanism operates competitively, with the target lemma reaching an activation threshold faster than competing lemmas (Roelofs, 1992). This selection is influenced by word frequency, semantic neighbourhood density, and context.

When the necessary lemma is chosen, the final level, phoneme activation, is activated. This is particularly true if accessing the lexeme (encode of the phonological form of the word) is included (Meyer, 1996). Phonological units (phonemes) are accessed and arranged in the right order of time for speech preparing. This type of stage is subject to phonological neighbourhood constraints and may include cascading activation between higher and lower levels (Dell et al., 1997).

The sequence of these stages is not strictly linear and some of them may be taken in parallel chronologically, as seen by a combination of chronometric and neuroimaging studies, supporting a cascaded processing model of lexical access (Indefrey & Levelt, 2004; Strijkers & Costa, 2011). Conceptual activation, lemma selection, and phoneme retrieval explain how individuals retrieve and produce words during naming tasks.

1.2 Naming tasks and their variants

Naming involves the retrieval and production of a word from a concept, object, or stimulus. It is a cognitive-linguistic process and is an important measure of lexical retrieval ability, and is often employed in experimental and neuropsychological research to measure language function per se and the integrity of the mental lexicon (Levelt, 1989; Goodglass & Kaplan, 1983). Standardised tests such as the Boston Naming Test (BNT) (Kaplan, Goodglass & Weintraub, 1983) and such as the Philadelphia Naming Test (Roach, Schwartz, Martin, Grewol and Brecher 1996) are frequently used while picture naming and word retrieval tasks are also included in more comprehensive language batteries such as the Western Aphasia Battery (WAB) and the Boston Diagnostic Aphasia Examination (BDAE).

Such tasks often require the retrieval of words in response to visual input, verbal cues or context, thereby activating different input modalities and processing routes, and can thus provide different aspects of information regarding activation and retrieval of words. To circumvent this problem, an Auditory Naming Test (ANT; Hamberger & Seidel, 2003) was developed as a supplementary measure that provides auditory-based semantic cues and, therefore, patterns like more naturalistic language processing requirements. This shift toward multimodal assessments highlights the importance of more diverse naming tasks to improve diagnostic sensitivity and ecological validity. There are several naming task variants, each tapping into slightly different cognitive mechanisms:

1.2.1 Confrontation Naming

In confrontation naming, people have to name a picture or object that is shown to them. It is the most common naming task used in neurocognitive studies. It involves

a bottom-up process that starts with visual input, followed by conceptual activation and finally lexical and phonological retrieval (Indefrey & Levelt, 2004).

The factors that affect Confrontation naming performance include demographic factors (e.g., age, education, gender) and linguistic factors (e.g., word frequency, age of acquisition). Words that are high frequency and early acquired are generally easier to retrieve and more accurate, while low frequency words add to the variability. In other neurological disorders such as Alzheimer's disease, temporal lobe epilepsy and aphasia, the errors made may also be quantitatively different, but not qualitatively different. Sometimes phonemic cues are given to facilitate retrieval, but their clinical utility is questionable. But there are also some drawbacks to confrontation naming.

Confrontation naming relies on visual processing, and therefore may not be a reliable measure of word retrieval in patients with visual agnosia, optic deficits, or other perceptual impairments. Furthermore, it does not encompass all the modality-specific retrieval difficulties that can occur in natural communication. To overcome these drawbacks, researchers (specifically cite) suggest using auditory-based tasks (such as the Auditory Naming Test) to measure lexical access using verbal prompts in addition to confrontation naming. This multimodal approach offers a more comprehensive understanding of naming performance and enhances diagnostic sensitivity.

The generative naming task (also known as verbal fluency) is widely used in both clinical and research settings to assess lexical retrieval, in addition to confrontation naming. This task asks participants to produce several exemplars for a given lexical category, either semantically (e.g., animals, fruits) or phonemically (e.g.,

words that start with the letter “F”). In addition to semantic organisation and lexical access, it engages executive functions, including cognitive flexibility, strategic search and working memory (Henry & Crawford, 2004).

Semantic fluency tends to rely more heavily on temporal lobe structures associated with semantic memory. In contrast, phonemic fluency recruits left frontal regions, particularly the dorsolateral prefrontal cortex, associated with goal-directed search and self-monitoring (Baldo et al., 2006).

Together, generative naming and confrontation naming tasks provide complementary insights into lexical retrieval while generative naming task captures strategic, self-generated word search, a confrontation naming task on the other hand directly measures stimulus-driven naming ability, allowing clinicians to examine both spontaneous lexical access and cue-dependent retrieval.

The Boston Naming Test (BNT) is a confrontation naming test developed by Kaplan, Goodglass, and Weintraub (1983), is a widely used confrontation-naming measure designed to assess lexical retrieval abilities in individuals with language impairments. It consists of black-and-white line drawings arranged in increasing order of difficulty, beginning with high-frequency, familiar objects and progressing to low-frequency, less common items. During administration, the individual is asked to name each picture spontaneously; if unsuccessful, semantic or phonemic cues may be provided to determine whether retrieval improves with support. The test is particularly sensitive to naming deficits associated with aphasia, dementia, and other neurological conditions, as it allows clinicians to examine error types, cue responsiveness, and patterns of lexical access breakdown. Its structured format and graded item difficulty

make it a valuable tool for identifying mild to severe anomia and for monitoring changes in word-finding abilities over time.

Similarly, the Western Aphasia Battery (WAB), developed by Kertesz (1982), is a comprehensive assessment tool used to identify the presence, type, and severity of aphasia in adults with acquired neurological disorders. It evaluates four core language domains, such as, spontaneous speech, auditory comprehension, repetition, and naming using structured tasks that capture both the qualitative and quantitative aspects of communication. The WAB generates an Aphasia Quotient (AQ), which reflects functional language ability and assists in classifying aphasia subtypes such as Broca's, Wernicke's, conduction, anomic, and global aphasia. Additional subtests assess reading, writing, praxis, and constructional abilities, providing a broader understanding of cognitive-linguistic functioning. Its standardized scoring system, diagnostic clarity, and ability to track progress across therapy make the WAB a widely relied-upon instrument in clinical and research settings.

Also, the Boston Diagnostic Aphasia Examination (BDAE), developed by Goodglass and Kaplan (1983), is a comprehensive battery designed to assess language impairments following brain injury and to classify aphasia into traditional Boston syndromes. It evaluates a wide range of linguistic domains, including conversational and expository speech, auditory comprehension, oral expression, reading, and writing, using tasks that capture both qualitative error patterns and quantitative performance. A key feature of the BDAE is its emphasis on detailed profile analysis, which helps distinguish among aphasia types such as Broca's, Wernicke's, conduction, anomic, and transcortical aphasias. The test also includes the Boston Naming Test and supplemental measures that provide deeper insight into lexical retrieval, syntax,

phonology, and paraphasic errors. Its structured yet descriptive approach allows clinicians to identify underlying processing deficits, guide treatment planning, and monitor changes over time, making the BDAE one of the most influential and widely used aphasia assessments.

1.2.2 Responsive Naming and Auditory Naming Task

Responsive naming involves naming based on a description or question without visual input. For example, when asked, “What do you drink from?”, the expected response would be “cup/Bottle/Tumbler.” This task directly taps into semantic memory and auditory comprehension, relying on verbal prompts to initiate conceptual search (Benson & Ardila, 1996). Responsive naming helps assess semantic access independent of visual processing demands; however, the response may not be specific compared to a confrontation naming task. In the confrontation naming task, a specific lemma node corresponding to the picture will be activated, while multiple lemma nodes can correspond to each target in the responsive naming task.

Responsive naming, while commonly used to probe semantic memory, also indirectly taps into verbal working memory and inferential reasoning, Verbal working memory refers to the temporary storage and manipulation of verbal or language-based information that is necessary for complex cognitive tasks such as comprehension, learning, and reasoning (Baddeley, 2000), on the other hand, Inferential reasoning is the cognitive ability to derive logical conclusions and make predictions from available information, even when that information is incomplete or implied as stated by (Graesser et al., 1994) Graesser, Singer, & Trabasso (1994, p. 371) in their study, since individuals must maintain the verbal cue in memory while searching their semantic network for reasonable mental representation of word in mental lexicon (Nickels,

2002; Saffran, 1997). These demands increase in complexity when the prompt/stimulus becomes more abstract or when multiple answers could be appropriate, reflecting a diffuse activation of lemma-level representations (Dell et al., 1997; Levelt, 1999; Meyer, 1996; Roelofs, 1992).

It is often considered a structured sub-type of responsive naming, presenting the naming cue exclusively in the auditory modality, typically using descriptive sentences or definitions (Hamberger & Seidel, 2003). This makes auditory naming more ecologically valid, simulating real-world language use, where referents are not presented as pictures but instead given spoken input (Gorno-Tempini et al., 2011). Additionally, the short presentation period of auditory input introduces additional processing load on phonological loop components of working memory (Baddeley, 2000; Saffran, 1997). Both tasks allow for an examination of modality-specific naming. From a neuroanatomical perspective, auditory naming performance has been associated with the integrity of the posterior superior temporal gyrus, planum temporale, and middle temporal gyrus regions implicated in processing and integrating complex auditory stimuli (Hamberger & Seidel, 2003; Henry et al., 2016). In clinical settings, auditory naming serves as an effective measure of:

- Lexical retrieval under temporal constraints (Saffran, 1997)
- Differentiation between semantic vs. phonological naming deficits (Gorno-Tempini et al., 2011; Nickels, 2002).
- Cross-modality comparison with visual naming tasks (Hamberger & Seidel, 2003).

Randolph et al. (1999) investigated how demographic and linguistic factors influence confrontation naming using a large sample of 180 neurologically healthy adults aged between 20 and 80 years. Participants completed the Boston Naming

Test (BNT), and their performance was analyzed in relation to variables such as age, education, and word frequency. The researchers found that naming accuracy decreased with age and lower education levels, particularly for low-frequency items. They concluded that confrontation naming relies not only on lexical access but also on cognitive reserve and lexical familiarity, emphasizing the role of experiential and linguistic variables in word retrieval.

Hamberger and Seidel (2003) created the Auditory Naming Test (ANT) and gave it to 25 patients with left or right temporal lobe epilepsy and 25 healthy controls. For this task, participants were given short verbal instructions, e.g., “What do you write with?” and asked to answer with one word (e.g., “pen”). The performance in this auditory naming task was compared to that in a visual confrontation naming task. The results revealed that patients with left temporal lobe epilepsy were significantly impaired on the auditory task, suggesting that anterior temporal regions involved in semantic cue integration play a crucial role in auditory lexical access.

On this basis, Hamberger et al. (2007) directly investigated brain regions involved in the different naming modalities in nine patients undergoing epilepsy surgery during the procedure. In awake craniotomy, patients were instructed to identify visual objects and answer to verbal descriptions of objects as specific areas of the cortex were briefly stimulated. Performance disruptions were noted and localized to cortical sites. The study found that anterior temporal sites were more likely to be disrupted by stimulation of the auditory naming system, while posterior temporal sites were more likely to be disrupted by stimulation of the visual confrontation naming system. This offered strong causal support for the existence of separate but related networks for naming by modality.

In a similar study, Tomaszewski Farias et al. (2005) used fMRI to study 30 healthy right-handed adults while they were performing visual and auditory naming tasks. Participants were presented with pictures or verbal definitions of the target items in an alternating manner and were asked to identify the target items. They discovered that the anterior temporal lobe and inferior frontal gyrus were more strongly activated during auditory naming, whereas posterior temporal and occipital regions were more strongly activated during visual naming. The study showed that the transient nature of spoken input necessitates more semantic integration and working memory involvement in auditory naming.

The generative naming paradigm has also been used to tap the lexical retrieval from a cognitive-control perspective. Henry and Crawford (2004) evaluated 60 patients (30 with frontal lobe damage and 30 healthy controls) on semantic (e.g., “animals”) and phonemic (e.g., “words beginning with F”) fluency tasks. In 60 seconds, participants came up with as many relevant words as they could. The researchers examined patterns of clustering (semantic grouping) and switching (shifting between subcategories). Results revealed that the naming processes were related to both semantic memory (primarily to the temporal lobe structures) and executive control (primarily to frontal-executive regions), indicating that naming relies on both semantic memory and executive control.

Last, Gorno-Tempini et al. (2011) investigated 12 patients with semantic dementia to compare auditory and visual naming. The participants were required to identify objects in pictures and to generate words when presented with verbal definitions. MRI results were compared to performance on both tasks. Patients demonstrated more deficits on auditory naming than visual naming, and this was

correlated with more anterior temporal lobe atrophy in the left hemisphere. This confirmed the importance of this area in the process of associating auditory and semantic information in word retrieval.

All of these studies indicate that confrontation, auditory, responsive, and generative naming tasks recruit similar, but different, cognitive and neural systems. The naming of confrontation depends on visual recognition and lexical retrieval, whereas the naming of auditory and responsive depends on verbal comprehension, temporal sequencing, and semantic inference. Generative naming introduces the executive control and strategic search elements. The model of lexical retrieval that is consistently supported by lesion, neuroimaging and stimulation studies is a modality-specific but interactive model that suggests that the neural pathways and cognitive strategies involved in naming are determined by the demands of the task and the modality of the input.

Perret and Bonin (2019) conducted a study that had a strong impact on the field, examining how the age of acquisition, imageability, and visual complexity of pictures affect picture naming in French-speaking adults. Their results showed that early-acquired and highly imageable words were retrieved more efficiently, thus supporting the traditional picture naming paradigm and the fact that psycholinguistic factors that affect lexical organization have a strong influence on it. In line with this, Rofes et al. (2018) conducted a large cohort study of healthy adults to compare the performance of object naming and action naming tasks, and found that action naming resulted in slower response times and more semantic variability than object naming. The results indicated that verb retrieval requires extra cognitive resources, probably

due to the fact that actions are more complex than nouns in terms of their semantic and syntactic representation.

On this basis, Mätzig et al. (2009) examined different types of naming tasks, such as picture naming, auditory naming, and definition-based naming. They found that all variants rely on lexical retrieval, but that the modality of input has a strong effect on the processing pathways: auditory cues involve phonological-to-lexical mapping, while visual cues involve conceptual-to-lexical activation. To this end, Perret et al. (2013) reported that reaction times were longer for auditory naming than for picture naming, indicating that the lack of direct visual cues increases the cognitive load involved in lexical search.

In contrast, Khwaileh et al. (2018) studied referential (picture-based) and inferential (definition-based) naming in monolingual adults. They found that their study showed that there were different temporal activation patterns, with inferential naming activating wider semantic networks before lexical selection. This indicates that naming tasks differ not only in perceptual modality but also in terms of the semantic integration which is required for successful retrieval. Likewise, Perret and Bonin (2018) showed that the presence of contextual cues in naming tasks improves the retrieval fluency, thus emphasizing the role of task context in influencing access efficiency.

The importance of temporal structure as an important variant in naming paradigms has also been studied. Georgiou et al. (2013) conducted a study that compared isolated naming tasks with rapid serial naming (RSN) tasks in school-aged children. Rapid Serial Naming (RSN) is the ability to name a rapid succession of familiar visual stimuli (letters, numbers, colors, objects) quickly and accurately. It is

a measure of the efficiency of visual recognition, lexical access and phonological retrieval, and is therefore a critical measure of processing speed and automaticity in word retrieval. Because of the relationship between RSN and reading fluency and naming, difficulties with RSN are often used to understand problems with reading fluency and naming. They determined that the performance of RSN was a better predictor of reading fluency. This difference is due to the fact that serial naming tasks measure the speed of retrieval and processing of information in an automatic way, while single item naming measures deliberate lexical search. The study therefore validated the hypothesis that temporal dynamics of naming tasks have a significant impact on cognitive demands and performance outcomes.

Abreu-Mendoza et al. (2020) conducted a study on cross-modal naming by analysing visual, auditory, and semantic cue-based tasks among bilingual adults. Their findings indicated modality-specific advantages contingent on language dominance, indicating that naming performance is not consistent across the task variants but rather interacts dynamically with linguistic experience and processing demands. These findings collectively demonstrated that naming tasks encompass a wide methodological spectrum, where each variant emphasises a distinct aspect of lexical retrieval, such as modality-specific access, semantic integration, or retrieval speed.

1.3 Utility and variables influencing auditory naming tasks

Apart from lexical frequency and age of acquisition, semantic category and word concreteness are important factors that affect the auditory naming performance. Concrete nouns (e.g., “apple”) are, for example, named more accurately and quickly than abstract nouns (e.g., “justice”) because they have richer sensory and perceptual representations that can be retrieved via multiple semantic pathways. (Hamberger &

Seidel, 2003) Likewise, Shao et al. (2014) showed that the familiarity of the semantic category influences lexical access speed, with objects and actions that are more familiar in daily life (e.g., body parts, household items) being accessed more quickly because they have stronger semantic connections.

Lexical neighborhood density (LND), defined as the number of phonologically similar words, has also been found to be a factor in naming latency and accuracy. Vitevitch and Luce (1998) found that words from dense phonological neighborhoods are retrieved more efficiently due to facilitative activation spreading across interconnected lexical nodes that aids in the resolution of competition during auditory word retrieval. This finding highlights the importance of phonological competition and lexical inhibition in auditory naming, similar to what is found in speech perception and production.

Auditory naming performance is also modulated by cognitive-linguistic variables like working memory capacity and attention. Martin and Saffran (1997) pointed out that the auditory naming task demands temporary maintenance of verbal input (the spoken definition) while searching for a lexical match, which places demands on the phonological loop component of working memory. This can result in slower reaction times or retrieval errors, particularly when using complex or multiword auditory cues, due to working memory or attentional control deficits. Likewise, Baldo et al. (2013) reported that patients with left prefrontal lesions had deficits in auditory naming, and that the integrity of executive control mechanisms that guide the selection of competing lexical alternatives was associated with successful retrieval.

There have also been consistent differences found between age groups. Bowles et al. (2016) found that older adults have slower response times for auditory naming, but accuracy is preserved, suggesting age-related changes in processing speed and semantic activation thresholds. The results are consistent with the Transmission Deficit Hypothesis (Burke & Shafto, 2008) which suggests that a failure in the transmission of information within the lexical-semantic network is responsible for word-finding problems in aging, especially when the information is presented only auditorily, without visual presentation.

Neuroanatomically, Hamberger et al. (2007) showed that lesions or stimulation to the anterior temporal lobe affected auditory naming but not visual confrontation naming, supporting modality-specific localization. Malow et al. (1996) also reported that stimulation of the superior temporal gyrus impaired comprehension-based naming, highlighting the importance of temporal lobe integration in the mapping of auditory input to semantic representations. Furthermore, Damasio et al. (2004) proposed that the posterior temporal areas are involved in visual naming, whereas the anterior and middle temporal cortices are important for auditory naming, especially for combining semantic and phonological codes.

The organization of lexical and semantic properties of a language also influences the efficiency of auditory naming as revealed by cross-linguistic studies. Shao et al. (2014) and Bürki & Madec (2023) noted that there is more variability in retrieval latency in morphologically rich languages, indicating that languages with complex inflectional morphology place more cognitive demands during auditory naming. This supports the hypothesis that linguistic typology and cultural familiarity

influence naming performance and should be taken into account when developing and standardizing tests.

In general, these results show that a complex interplay of linguistic (lexical frequency, semantic category, neighborhood density), cognitive (working memory, attention, executive control), and neuroanatomical (temporal and frontal cortical regions) factors influences auditory naming. It is important to be aware of these influences when designing valid and culturally appropriate auditory naming tasks and when interpreting differences in performance among populations.

1.4 Research gaps related to the Auditory Naming Test

Despite the valuable contributions of Hamberger and colleagues in developing and validating the Auditory Naming Test (ANT), several research gaps remain unaddressed. Firstly, the normative data provided by Hamberger and Seidel (2003) were limited to English-speaking adults from relatively homogeneous educational and cultural backgrounds, restricting the cross-linguistic and cross-cultural generalizability of the test. As lexical retrieval is influenced by linguistic typology and sociocultural experience (Bialystok et al., 2004), auditory naming tools developed solely in Western contexts may not reflect the lexical access demands of multilingual populations. Languages such as Hindi, Kannada, and Tamil, spoken by million of South Asians, exhibit few unique morpho-syntactic features and lexical frequency distributions that can affect the auditory word retrieval patterns (Annamalai, 2001).

Secondly, the ANT has been explored primarily in English, a morphologically impoverished language. This leaves questions about its applicability in morphologically rich and agglutinative languages, where lexical retrieval may involve more complex morphological parsing (Rastle et al., 2004). Cross-linguistic

psycholinguistic studies suggest that language-specific lexical and phonological structures significantly influence naming performance (Szekely et al., 2004).

In contrast, adjunct ANT offers a modality-specific assessment of auditory lexical access; comparative studies do not examine its sensitivity, item difficulty, and reaction time profiles relative to visual naming tasks (Barry et al., 1997; Nickels, 2002). Such comparisons are essential to determine whether auditory and visual naming engage parallel or distinct processing pathways, especially in languages with phonologically opaque orthographies.

Finally, current auditory naming assessments are scarce in many non-Western languages. Given the cognitive load and working memory demands unique to auditory naming (Levelt, 1999), designing linguistically and culturally adapted tools is crucial. Without such tools, it is challenging to study modality-specific lexical access or to map developmental trends in auditory naming across age groups and languages (Bates et al., 2001; Kohnert, 2004).

Chapter 3

Method

The Hindi Auditory Naming Test (HANT) was developed in three phases. In the first phase, 30 stimulus questions were created in Hindi to encourage participants to respond with a single word, either a concrete or an abstract noun, based on the type of stimulus question presented. These questions were designed to sound natural and similar to how an individual speaks in everyday situations (in a colloquial manner).

In the second phase, the stimuli underwent content validation by 5 experts in speech-language pathology. Experts evaluated each question for Clarity, Ambiguity, Relevance, using a 3 point Likert scale, and the Item-wise Content Validity Index (I-CVI) was calculated. Additionally, Familiarity, and Appropriateness ratings for the target words were gathered to ensure cultural and linguistic appropriateness for Hindi-speaking adults. This process guided the refinement and final selection of test items.

In the third phase, the validated questions were administered to 60 neurotypical adults across different age groups. The auditory stimuli were presented verbally, and participants were instructed to respond verbally. All responses were audio-recorded for later analysis. The data were examined to assess naming accuracy and to observe performance trends across age and gender groups, helping to determine the test's fairness and suitability for broader use.

The study incorporated qualitative analysis of participant responses to examine the nature and patterns of errors, thereby contributing to a more comprehensive understanding of lexical retrieval processes and enhancing the clinical applicability of

the tool. In addition, to evaluate the stability and consistency of the measure over time, 10% of the collected data were subjected to test–retest reliability analysis.

3.1 Selection of Target Stimulus

The Hindi Auditory Naming Test (HANT) stimuli in Hindi were developed using a linguistically grounded approach to ensure ecological validity and clinical relevance. A mix of frequent and infrequent Hindi nouns was selected to maintain a gradual complexity gradient, based on the intended target word (as listed by the investigator). The final stimulus set included a balanced mix of high- and low-frequency words across semantically meaningful categories such as animate/inanimate entities, body parts, tools, natural elements, and other lexical categories.

Key psycholinguistic variables like word frequency were controlled to minimise variability. Each word was embedded in a short, semantically neutral sentence (7–8 words) simulating natural spoken Hindi. These sentences varied in syntactic structure and word position (subject, object, verb) to reflect natural language use. Recordings were made by a native Hindi speaker using clear articulation and natural prosody.

Each word's prototype question was created based on dictionary definitions and contextual usage. Experienced speech-language pathologists reviewed these for relevance, clarity, ambiguity, familiarity and appropriateness in, eliciting the intended word. Cultural and contextual appropriateness was ensured by prioritising items relevant to daily Indian life, while excluding culturally unfamiliar or ambiguous terms.

Expert review and content validation refined items that passed these criteria, leading to a final set of 30 auditory stimulus questions.

The selection of target stimuli for the proposed Hindi Auditory Naming Test (HANT) was informed by a set of psycholinguistic, cultural, and linguistic parameters to ensure clinical relevance and ecological validity. The following criteria were systematically applied during the stimulus development process:

a) **Lexical Frequency and Familiarity**

Stimuli were selected based on their frequency of occurrence in spoken Hindi. Stimuli included both highly frequent words, which were easily recognised in everyday communication to minimise cognitive load (Hamberger & Seidel, 2003), and less frequently used words, which broadened the lexical difficulty range and helped prevent ceiling effects, thus enhancing the test's ability to discriminate among different levels of lexical retrieval.

b) **Complexity Gradient**

The complexity gradient was achieved by varying the target noun characteristics and the structure of the stimulus questions. Target words differed in frequency (high vs. low) and concreteness (concrete vs. abstract), as well as morphological complexity (simple vs. inflected). In addition, stimulus questions varied in syntactic structure and positioning of the target word (e.g., subject, object) to simulate natural linguistic contexts and increase processing demands. This dual-layered approach allowed for a more graded assessment of lexical retrieval across increasing levels of linguistic complexity.

c) **Cultural and Contextual Relevance.**

All stimuli were selected carefully, considering their cultural appropriateness and

functional relevance to Indian speakers. Items contextually grounded in typical daily life activities were prioritised, and culturally incongruent or unfamiliar terms were excluded to maintain ecological validity.

3.2 Stimulus Validation and Response Analysis Parameters

The stimulus questions and target words were analyzed to determine the linguistic appropriateness and the diagnostic value of each item in the Auditory Naming Test. Each question was assessed for its relevance (i.e., how directly it elicited the intended word) and clarity (i.e., how easily the target population could understand it), Ambiguity (i.e The item is unclear and can be interpreted in more than one way), Familiarity and appropriateness (i.e., the response accurately matches what the test intends to measure). Likewise, the familiarity of the target words was assessed in terms of cultural and lexical use and the difficulty level was assessed in terms of phonological and linguistic complexity. After pilot administration, an item-wise analysis was conducted to look for variability in the responses of the participants and consistency across age and gender groups. This multi-level validation enabled the test to be clarified, fair and representative.

Content validation was done in two phases to ensure that each test item of the Hindi Auditory Naming Test (HANT) was linguistically suitable and functionally relevant. The content validity measures were modified from the Manual of Non-Fluent Aphasia Therapy (Goswami et al., 2012).

In the primary stage, five speech-language pathology experts were asked to evaluate only the stimulus questions (not the target responses). A 3-point Likert scale was used for the ratings, 3 = Highly Relevant, 2 = Somewhat Relevant, 1 = Not

Relevant; 3 = Highly Clear, 2 = Somewhat Clear, 1 = Not Clear; 3 = Very Ambiguous, 2 = Somewhat Ambiguous, 1 = Not Ambiguous; and so on. An Item-wise Content Validity Index (I-CVI) was determined and the items with I-CVI 0.78 or above were retained, the other items were modified or eliminated based on the experts' feedback.

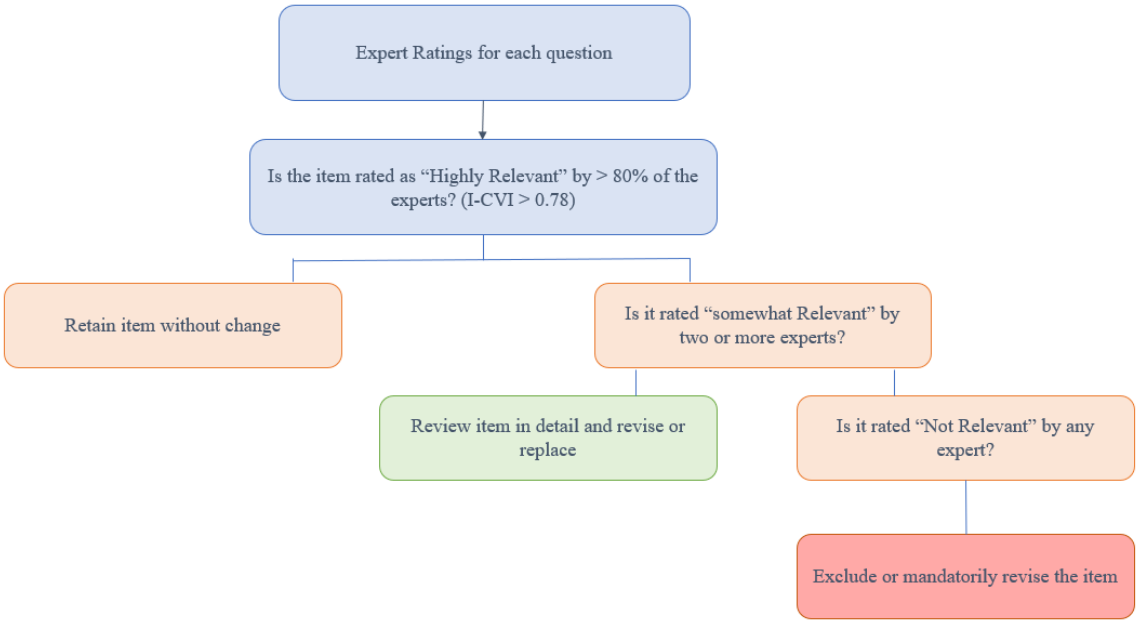
The same experts also rated the target words at this stage, independently, as "relevant" and "highly relevant." This same group rated the words for familiarity and appropriateness. The experts assessed the familiarity of each word and its appropriateness in relation to a typical Hindi speaking adult on a 3-point scale, Very familiar, Somewhat familiar, Less familiar and Highly appropriate, Somewhat appropriate, Not appropriate. This process enabled the identification of words which were too uncommon or culturally specific to be included. This was neither required for content validation but was a step taken to help refine the final test set and make it more culturally and linguistically relevant. These procedures helped guarantee that the final stimuli were meaningful, clear, and suitable for lexico-semantic responding in the target group. The Item-wise Content Validity Index (I-CVI) was obtained for each item according to the number of experts rating it 3 or 4. Those that received an I-CVI score of 0.78 or above (that is at least 4 out of 5 experts agreed) were kept (Lynn, 1986).

1. Items rated as "Not Relevant" by any expert were mandatorily revised, replaced, or skipped.
2. Items marked as "Somewhat Relevant" by two or more experts underwent optional revision, guided by qualitative suggestions.
3. Stimuli rated as "Highly Relevant" by all were retained unless improvement was specifically recommended.

To ensure cognitive-linguistic appropriateness, items rated as “Somewhat Relevant” by multiple experts were not immediately discarded. Instead, such items were closely reviewed and, where possible, paraphrased or restructured to improve clarity and alignment with the intended lexical target while preserving their clinical and linguistic value.

Figure 3.1

Content Validation Decision Process for Item Retention, Revision, and Exclusion



3.3 Participant Details

A total of 60 participants were included in the study. Participants were selected using convenience sampling and were divided into three age-based groups with equal representation, comprising 30 individuals. The age ranges for the groups were as follows: Group I included individuals aged 18 to 35 years, Group II included those aged 36 to 50 years, and Group III included individuals aged 51 to 65 years. Both males and females were considered for inclusion; however, the gender distribution was determined based on the availability of participants within each age group.

Table 3.1

Participants details of Group I

S. No.	Age	Gender	Education	Occupation	Hindi Proficiency
1	27	F	Graduate	Student	Native-like
2	23	F	Graduate	Student	Native-like
3	28	F	Graduate	Student	Native
4	22	F	Graduate	Student	Native
5	23	F	Graduate	Student	Native-like
6	23	F	Graduate	Student	Native
7	25	F	Graduate	Student	Native

8	28	F	Graduate	Student	Native
9	28	F	Graduate	Student	Native
10	28	F	Graduate	Student	Native
11	25	F	Graduate	Student	Native
12	23	F	Graduate	Student	Native-like
13	24	F	Graduate	Student	Native-like
14	23	F	Graduate	Student	Native-like
15	24	M	Graduate	Student	Native
16	24	F	Higher Secondary	Student	Native
17	21	F	Higher Secondary	Student	Native
18	21	F	Graduate	Student	Native
19	26	F	Graduate	Student	Native
20	24	F	Graduate	Student	Native

Table 3.1 shows 20 adult participants of Group I, within the age range of 18–35 years, with a mean age of approximately 24.5 years. This group consisted of female participants mostly, with only one male participant. Most individuals were students, and the majority had completed graduate-level education, and few have higher secondary qualifications. In this

group,14 participants demonstrated ‘native’ proficiency in Hindi, and 6 participants showed ‘native-like’ proficiency, which means that this group is linguistically competent to be a part of this study.

Table 3.2

Participants details of Group II

S no.	**Age**	**Gender**	**Education**	**Occupation**	**Hindi proficiency**
1	35	F	Graduate	home maker	Native
2	35	F	Graduate	home maker	native like
3	48	F	Graduate	home maker	Native
4	34	F	Graduate	corporate employee	Native
5	49	F	post graduate	home maker	Native
6	36	F	Graduate	physician	Native
7	37	F	Graduate	physician	native like
8	39	M	Graduate	corporate employee	native like
9	41	M	Graduate	physiotherapist	Native

10	41	M	post graduate	teacher	Native
11	44	M	Graduate	SLP	Native
12	37	M	Graduate	corporate employee	Native
13	36	M	post graduate	SLP	Native
14	38	M	Graduate	student	Native
15	36	F	Graduate	student	native like
16	35	F	Graduate	SLP	Native
17	36	F	Graduate	teacher	Native
18	34	M	Graduate	teacher	Native
19	46	M	post graduate	teacher	Native
20	49	M	post graduate	teacher	Native

Table 2.2 shows 20 adult participants of Group II, who fall within the age range of 34-50 years, with a mean age of approximately 39.3 years. This sample consisted of a more balanced gender distribution compared to Group I. Most individuals had completed graduate or post-graduate education. Widely, the group included homemakers, teachers, corporate employees, healthcare professionals (physicians/physiotherapists), and speech-language pathologists, which shows a diverse professional background. Regarding language proficiency, 15 participants

demonstrated ‘native’ proficiency in Hindi, while 5 participants showed ‘native-like’ proficiency, indicating that all participants were linguistically competent.

Table 2.3

Participants details of Group *III*

S no.	**Age**	**Gender**	**Education**	**Occupation**	**Hindi proficiency**
1	52	F	10th pass	homemaker	Native
2	53	F	8th pass	homemaker	Native
3	56	F	8th pass	homemaker	Native
4	58	F	12th pass	homemaker	Native
5	53	F	10th pass	homemaker	Native
6	61	M	graduate	teacher	Native
7	51	M	graduate	teacher	native like
8	51	M	12th pass	singer	Native
9	63	M	12th pass	singer	Native
10	62	M	12th `pass	retired	Native
11	51	M	12th pass	singer	Native

12	54	F	12th PASS	buisnessman	Native
13	55	F	12th pass	buisenessman	Native
14	52	F	10th pass	homemaker	Native
15	56	M	graduate	homemaker	Native
16	60	M	Higher secondary	retired	Native
17	62	F	10th pass	homemaker	Native
18	64	F	graduate	homemaker	Native
19	55	M	10th pass	Clerk at gov.	Native
20	58	M	graduate	Clerk	Native

Table 2.3 represents 20 adult participants of Group III, age range from 51–65 years and a mean age of approximately 56.5 years. This group included a slightly higher proportion of females, many, were homemakers. Participants had wider variety of educational backgrounds compared to previous groups, such as 8th pass to graduate-level qualifications. Their occupations included homemakers, teachers, singers, clerks, businessmen, and retired individuals,. Nineteen participants had ‘native’ proficiency in Hindi, while 1 participant exhibited ‘native-like’ proficiency, hence this group also had proficient hindi speakers.

3.3.1 Inclusion Criteria

The selection of participants was done according to certain inclusion criteria to guarantee the validity and appropriateness of the data collected. Only native-language speakers were considered. In the sociocultural context, few individuals were monolingual, and multilingual individuals were included if they were deemed to have adequate proficiency in the test language. To ensure accurate auditory processing of the stimuli, the languages known and the corresponding level of proficiency were recorded with the self-rated Language Experience and Proficiency Questionnaire (LEAP-Q). The participants were all with normal hearing abilities and no history of hearing loss was reported (Paurnaik et al., 2023). Hearing status was informally assessed to exclude peripheral auditory deficits, such as the Ling Six Sound Test. Also, participants were either normal or corrected vision. To ensure that participants have the cognitive and linguistic abilities needed to comprehend the task and produce a response, they must have at least 10-12 years of formal education. Finally, all participants had scores within the cut-off range of 25 or higher on the Montreal Cognitive Assessment, Hindi version (MoCA-H), to exclude any underlying cognitive deficits. All of these inclusion criteria together ensured a representative and cognitively intact sample that could be validly included in the auditory and visual naming tasks.

3.3.2 Exclusion Criterion

Participants with any neurological or psychiatric disorders, cognitive impairments, or severe anxiety and attention deficits were excluded from the study, as these conditions could have interfered with language processing and task performance.

3.4 Stimulus and Procedure

In lexical retrieval tasks, word frequency influenced response accuracy. High-frequency words were generally retrieved more easily and quickly than low-frequency words due to their frequent exposure and stronger representation in an individual's mental lexicon. Conversely, low-frequency words presented a greater challenge to lexical access and were therefore valuable for assessing the depth and flexibility of word retrieval processes. One variable that was reported to systematically affect both the speed and accuracy of confrontation naming was the frequency with which the name of a given object appeared in the lexicon. High-frequency items were typically named with greater speed and accuracy (Randolph et al., 1999).

The study was conducted on neurotypical adult participants as a preliminary phase to establish normative data and validate the stimulus material before its administration to clinical populations. However, testing clinical groups was not an objective of the present study and was planned for future research. Surveying a typically developing population allowed examination of natural lexical retrieval patterns and response tendencies without interference from language or cognitive impairments. This was essential for determining whether the designed auditory stimulus questions were appropriate,

comprehensible, and capable of reliably eliciting the intended grammatical constructions, such as nouns and verbs. Additionally, it facilitated assessment of how word frequency (frequent vs. infrequent) influenced lexical access under typical language processing conditions.

Complexity was controlled to balance simple and moderately complex sentence structures, avoiding overly abstract or syntactically demanding formulations that could influence lexical retrieval independently of word frequency. This prevented ceiling effects, where participants might perform at a maximum level on overly simple items, limiting the ability to observe variability in performance and diminishing the test's sensitivity. Flexibility was incorporated by ensuring that the stimulus questions allowed spontaneous yet appropriate responses, reflecting natural language use rather than forced or restricted answers. These parameters collectively ensured that the stimuli were functional, ecologically valid, and effective for examining lexical access patterns in the intended population.

3.4.1 Instructions

Before administering the test, each participant received standardised instructions regarding the nature of the task. They were informed that they would hear pre-recorded auditory questions presented via headphones and were required to respond as quickly and accurately as possible using a single word or a brief phrase. The following verbatim instruction was provided:

"I will now put the headphones on your ears. You will hear some recorded questions one by one. You must listen to them carefully and answer each question in one word. If needed, a question can be repeated once. If you still do not know the answer, say the most correct word."

मैं अभी आपको हेडफ़ोन कानों पर लगाऊँगी। इसमें एक-एक करके कुछ रिकॉर्ड किए हुए सवाल पूछे जाएँगे। आपको उन सवालों को ध्यान से सुनना है और सुनकर एक शब्द में

उसका जवाब बोलना है। अगर ज़रूरत हो तो एक सवाल को एक बार और दोहराया जा सकता है। अगर फिर भी आपको जवाब न आए तो जो सबसे सही समझ में आए, वही बोल देना है।

/maĩ abʱiː aːpkoː ɦeːɖpʰoːn kaːnoː̃ pər ləgaːũgiː | ismeː ek‿ek kərke kʊt͡ʃ rikɔːɖ kiːe hʊeː sʌwaːl pʊt͡ʃʰeː d͡ʒaːẽgeː | aːpkoː ʊn sʌwaːloː̃ koː ɖ̪ʱjaːn seː sʊnnaː hɛː ɔːr sʊnkar ek ʊʌrɖ̪ meː uskaː d͡ʒawaːb boːlnaː hɛː | əgər d͡ʒəruːraṯ̪ hoː toː ek sʌwaːl koː ek baːr ɔːr doːɦraːjaː d͡ʒaː sʌkt̪aː hɛː | əgər pʰir biː aːpkoː d͡ʒawaːb nə aːe toː d͡ʒoː sʌbseː sʌhɪː sʌmʤʱ meː aːe, ʊoːhiː boːl ɖ̪eːnaː hɛː |/

Participants were instructed to listen carefully and provide their response after the complete presentation of each item. A standardised prompt (e.g., “Please try to answer”) was given once if a participant did not respond within five seconds. No corrective feedback was provided during the test. Practice items were administered before the test to ensure the participant understood the task format.

3.4.2 Test Environment

The test was conducted in a quiet, well-lit room with minimal distractions to ensure the participant’s comfort and focus. A sound-treated room was preferred to reduce background noise and maintain audio clarity, especially during the pilot and primary data collection phases. High-quality headphones and a reliable recording device were used to present the auditory stimuli and capture participant responses. Each participant was seated in front of a laptop or audio playback device, and the examiner remained present to guide and monitor the session.

3.4.3 Presentation of Test Stimulus

All test items were pre-recorded by a native Hindi speaker using a clear and neutral voice. The speaker maintained a normal conversational pace to preserve naturalness. These recordings were processed and edited using audio editing software to ensure consistency in volume, pacing, and clarity. Each auditory stimulus was checked for familiarity, relevance to everyday communication, and flexibility in response. The stimulus questions contained a maximum word length of eight words and were balanced in grammatical form, word frequency, and content category. Care was taken to avoid dialectal variations or culturally unfamiliar references. The finalised stimuli were arranged in a fixed order and coded for easy identification during data collection and analysis.

3.4.4 Scoring and Interpretation

Each participant's response was evaluated based solely on accuracy, judged using a pre-defined scoring guide developed during the pilot study phase. Responses were categorised into four types: correct, acceptable synonym, ambiguous, and incorrect. A response was considered accurate if it matched the target word identified during the pilot study, particularly those produced by at least 80% of participants.

Acceptable synonyms included responses that did not precisely match the target word but preserved the intended semantic meaning and were commonly used in the relevant linguistic and cultural context. These synonyms were compiled during the pilot phase and finalised by a panel of language experts or clinicians to ensure consistency and minimise subjective bias. Both correct and acceptable synonym responses were awarded a full score of 1.

Responses that were conceptually related but too vague, broad, or imprecise to reflect the target meaning were marked as ambiguous and assigned a partial score of 0.5. Unrelated, semantically incorrect, or off-target responses were marked incorrect and scored 0. No response or explicit indications of uncertainty (e.g., "I don't know") were also assigned a score of 0.

To gain insights beyond accuracy, qualitative analysis of the responses was undertaken to examine the nature and types of errors. Further analyses were conducted to examine trends across lexical categories (nouns vs. verbs), word frequency (high vs. low frequency), and age groups. To ensure scoring consistency across raters and over time, inter-rater reliability and test-retest reliability were established. For this purpose, 10% of the data were rated by three independent raters apart from the investigator.

3.5 Analysis

Objective 1: Content Validation

To assess the content validity of the Auditory Naming Test (ANT), expert ratings were used to compute the Item-wise Content Validity Index (I-CVI). Each item was rated across three parameters familiarity, relevance, and flexibility using a 3-point Likert scale. The I-CVI for each item was calculated as the proportion of experts rating the item as either 3 (quite relevant) or 4 (highly appropriate). An I-CVI value of 0.78 or above was considered acceptable (Lynn, 1986), as per standard recommendations for 5-7 raters. Items falling below this threshold were revised or excluded based on qualitative expert feedback to ensure linguistic appropriateness, cultural relevance, and functional effectiveness for lexical retrieval assessment.

Objective 2: Field Testing

Data collected from 60 neurotypical adults were analysed in two phases.

First, the Shapiro–Wilk Test of Normality was used to examine the distribution of accuracy scores across different age and gender groups, determining whether parametric or non-parametric tests would be appropriate.

For group comparisons, if the data were normally distributed, a One-Way ANOVA was applied to analyse differences in naming performance across the three age groups. If the data were not normally distributed, the Kruskal–Wallis H test was used as a non-parametric alternative. When significant differences were found, post-hoc comparisons were conducted using the Mann–Whitney U test for pairwise analysis between age groups that was for Incorrect responses particularly. whereas in case where no significant difference was seen, no post hoc test was applied. The independent variables included age, while the dependent variable was naming accuracy. All statistical analyses were carried out using the Statistical Package for the Social Sciences (SPSS), version 26.0.

Objective 3: Reliability Analysis

To ensure scoring consistency, inter-rater reliability and test-retest reliability were established using Cohen's Kappa coefficient, which determined the level of agreement among raters. These reliability analyses confirmed the stability and reproducibility of the scoring process and validated whether the ANT effectively captured lexical retrieval abilities comparable to established naming measures.

CHAPTER 4

Results

The present study mainly aimed to design the content of the Auditory Naming Test in Hindi and validate it to ensure it is linguistically appropriate and culturally relevant, to conduct field testing on a controlled sample to examine performance variations across three age groups (18-30, 31-45, and 46-60) and to determine inter-rater and test–retest reliability by analysing 10% of the collected data. The scoring system for the test was decided to categorize responses into three parameters: correct, partially correct, and incorrect. This approach was considered to avoid limiting the responses to binary classification, and to take continuum of lexical retrieval performance into consideration. The correct responses reflected appropriate retrieval of the target word, whereas the incorrect responses indicated failure to retrieve the target word. Partially correct responses were included to show the responses that demonstrated incomplete retrieval, such as semantic substitutions, phonological errors or circumlocutions. This graded scoring pattern increased the sensitivity of the assessment by allowing identification of subtle deficits in lexical access.

4.1 Objective 1: Content validation

Content validity was established using both Item level content validity Index (I-CVI) and Scale-level content validity index to establish a quantitative evaluation of the test items. The I-CVI was used to assess the relevance of each individual item based on expert ratings, this allowed identification of items that may require some modification or deletion. On the other hand, the Scale level content validity index (S-CVI) was calculated to check universal agreement over the validity of the contents of the test, which reflects the degree to which the scale adequately represents the

construct of the items. The combination of these two parameters is recommended in instrument development as it ensures both item specific precision and overall scale adequacy to strengthen the validity of the tool (Polit & Beck, 2006). Following the development of the content for the Hindi Auditory Naming Test (HANT), a total of 31 auditory stimuli were subjected to content validation, which is considered as an important psychometric criteria when developing a test material in literature. A structured content validation feedback questionnaire was developed with the help of Manual of Adult: Non Fluent Aphasia Therapy in Kannada by Goswami et al. in 2012. The questionnaire contained 31 stimuli to be rated upon 5 parameters that are; Clarity, Relevance, Familiarity and Ambiguity. using a 3-point Likert scale. Content validation was performed by five Speech-Language Pathologists (SLPs) with an average clinical experience exceeding seven years. The questionnaire was emailed to each validator individually. All raters reported to have native or near-native proficiency in Hindi, which was assessed through a self-rating scale ranging from 0 to 10; ratings of 8 and above were considered highly proficient. Details of the validators are presented below.

Table 4.1

Demographic and proficiency details of raters.

S. No.	Years of Formal Education in Hindi	Designation	Hindi Proficiency
1	12 years	Assistant Professor, Speech-Language Pathologist	Native-like
2	12 years	Assistant Professor, Speech-Language Pathologist	Native-like
3	12 years	Assistant Speech-Language Pathologist	Native-like
4	12 years	Speech-Language Pathologist	Native-like
5	12 years	Speech-Language Pathologist	Native-like

The data was then assessed to check universal agreement and hence the content validity index was calculated. Based on the obtained ratings, the Item-level Content Validity Index (I-CVI) and Scale-level Content Validity Index (S-CVI) were calculated to determine the adequacy of the developed stimuli.

4.1.1 Item-Level Content Validity Index (I-CVI)

The Item-Level Content Validity Index (I-CVI) was used to quantify the degree of agreement among expert validators regarding the relevance and representativeness of each test item. It provides an objective, widely accepted method to identify items that adequately reflect the construct being measured and to point out

the items requiring revision or elimination, thereby strengthening the content validity of the tool.

The I-CVI analysis revealed high level of agreement among expert raters for the majority of the items. Most of the stimuli (n = 27) achieved an I-CVI of 1.00, indicating universal agreement regarding their adequacy.

A small number of items showed slightly lower agreement:

- Item no.3: I-CVI = 0.80
- Item no. 29: I-CVI = 0.80
- Item no. 30: I-CVI = 0.80
- Item no. 31: I-CVI = 0.60

Items with I-CVI values of 0.80 were still above the commonly accepted cut-off of 0.78 for five raters, suggesting acceptable content validity. However, Item 31, with an I-CVI of 0.60, fell below the acceptable threshold, indicating comparatively lower expert agreement and the need for review or modification.

Overall, the findings demonstrated that the majority of the developed auditory stimuli holds strong content validity.

Table 4.2

Results of item validation done by nine expert raters for concrete stimuli

Stimulus	Clarity	Relevance	Familiarity	Ambiguity	Cultural Appropriateness	I-CVI / UA
1	1.00	1.00	1.00	1.00	1.00	1.00
2	1.00	1.00	1.00	1.00	1.00	1.00
3	1.00	0.80	1.00	0.80	1.00	0.80
4	1.00	1.00	1.00	1.00	1.00	1.00
5	1.00	1.00	1.00	1.00	1.00	1.00
6	1.00	1.00	1.00	0.80	1.00	1.00
7	1.00	1.00	1.00	1.00	1.00	1.00
8	1.00	1.00	1.00	1.00	1.00	1.00
9	1.00	1.00	1.00	1.00	1.00	1.00
10	1.00	1.00	1.00	1.00	1.00	1.00
11	1.00	1.00	1.00	1.00	1.00	1.00
12	1.00	1.00	1.00	1.00	1.00	1.00
13	1.00	1.00	1.00	1.00	1.00	1.00

Table 4.3

Results of item validation done by nine expert raters for abstract stimuli

Stimulus No.	Clarity	Relevance	Familiarity	Ambiguity	Cultural Appropriateness	I-CVI
14	0.80	1.00	1.00	1.00	1.00	1.00
15	1.00	0.80	1.00	1.00	1.00	1.00
16	1.00	1.00	1.00	1.00	1.00	1.00
17	1.00	1.00	1.00	1.00	1.00	1.00
18	1.00	1.00	1.00	1.00	1.00	1.00
19	1.00	1.00	1.00	1.00	1.00	1.00
20	1.00	1.00	1.00	1.00	1.00	1.00
21	1.00	1.00	1.00	1.00	1.00	1.00
22	1.00	1.00	1.00	1.00	1.00	1.00
23	0.80	1.00	1.00	1.00	1.00	1.00
24	1.00	1.00	1.00	1.00	1.00	1.00
25	1.00	1.00	1.00	0.80	1.00	1.00
26	1.00	1.00	1.00	1.00	1.00	1.00
27	1.00	1.00	1.00	1.00	1.00	1.00
28	1.00	1.00	1.00	1.00	1.00	1.00
29	1.00	1.00	1.00	1.00	1.00	0.80
30	1.00	1.00	1.00	1.00	1.00	0.80
31	1.00	1.00	1.00	1.00	1.00	0.60

4.1.2 Scale-Level Content Validity Index (S-CVI)

The Scale-level Content Validity Index (S-CVI) was calculated separately for concrete and abstract stimuli to ensure adequate representation of content across these semantic domains. Since auditory naming performance may vary across these domains,the domain-wise analysis was carried out to verify that both concrete and abstract response eliciting stimuli independently demonstrated satisfactory level of content validity. This step was important to ensure balanced stimulus quality and strengthened the overall robustness of the developed Auditory Naming Test.

The scale-level analysis further supported the adequacy of the test content. The calculated S-CVI for the concrete response eliciting stimulus set was 0.988, indicating excellent overall content validity of the scale. Whereas, S-CVI value for abstract response eliciting stimuli found to be 0.943 which also fell well above the recommended benchmark of 0.90, further confirming strong expert agreement at the scale level.

These values suggest that the Auditory Naming Test demonstrates high overall content representativeness, clarity, and relevance as judged by experienced SLPs.

Table 4.4

Scale content validity index (S-CVI)

Domain	Number of Items	S-CVI
Concrete stimuli	17	0.988
Abstract stimuli	14	0.943

Overall, the findings indicated that the developed stimuli holds good relevance and were appropriate for inclusion in the Auditory Naming Test. However, one stimulus (Item 31), which obtained the lowest item-wise Content Validity Index (I-CVI), was excluded from the final stimulus set. Consequently, the finalized version of the test comprised 30 validated stimuli.

4.2 Objective 2:

The second objective of the study was to conduct field testing on a controlled sample to examine performance variations across three age groups (18-30, 31-45, and 46-60 years) and to assess the performance variations across the three age groups.

The participant's responses were categorized into the number of Correct, Incorrect and Partially correct, along with this the raw scores were also calculated. Prior to inferential analysis, the distribution of the data was examined using the Shapiro-Wilk test to check if the data follows normal or non normal distribution. The Shapiro-Wilk test results showed that the distribution of correct responses and raw scores did not significantly deviate from normality across age groups i.e $p > 0.05$. However, partial responses and incorrect responses showed significant deviation from normality in one or more age groups ($p < 0.05$). Hence, parametric statistics were used for normally distributed variables like number of correct responses and raw scores, whereas non-parametric tests were applied for variables that did not follow normality i.e number of incorrect and partial responses.

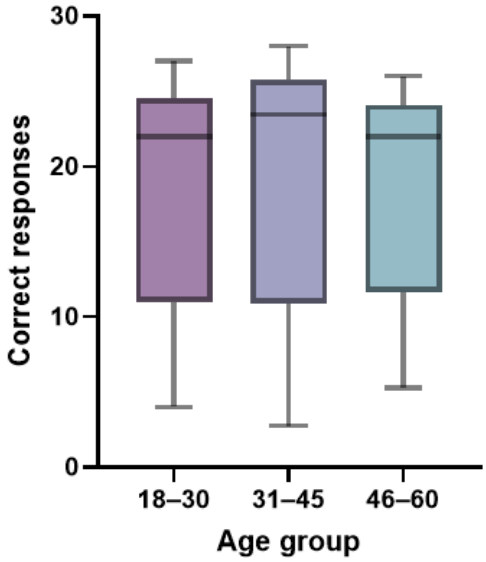
Table 4.5

Descriptive Statistics across Age Groups

Response Variable	Age Group (Years)	N	Mean	SD	Median	IQR
Correct Responses	18–30	20	22.10	2.38	22.00	4.00
	31–45	20	23.45	2.26	23.50	2.75
	46–60	20	22.20	2.53	22.00	5.25
Raw Score	18–30	20	23.78	2.03	24.00	3.25
	31–45	20	24.58	1.76	24.50	2.63
	46–60	20	23.65	2.22	23.75	3.25
Partial Responses	18–30	20	3.35	1.95	3.00	2.75
	31–45	20	2.25	1.37	2.00	0.00
	46–60	20	2.90	1.55	2.50	2.00
Incorrect Responses	18–30	20	1.40	0.88	2.00	1.00
	31–45	20	0.65	0.67	1.00	1.00
	46–60	20	1.80	1.06	2.00	1.75

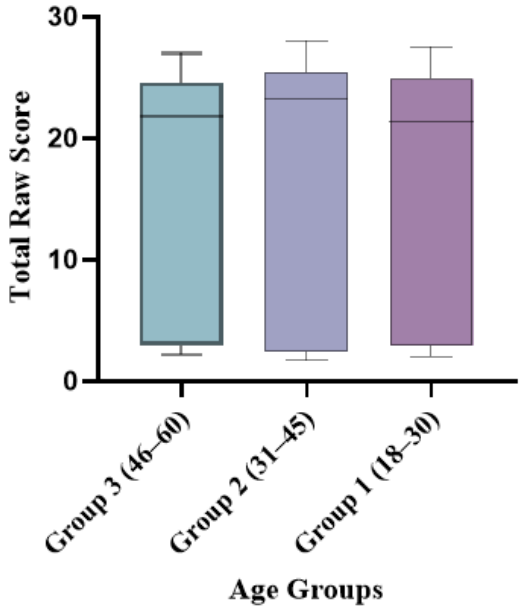
A comparison of the descriptive statistics across the three age groups was carried out which revealed noticeable trends in the measures of central tendency as well as dispersion.

Figure 4.1 *Box-and-Whisker Plot Showing Distribution of Correct Responses Across Groups*



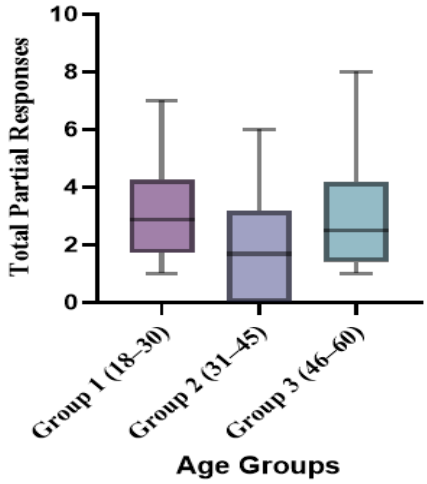
To study and analyze the correct responses, Group 2 demonstrated the highest mean score (Mean = 23.45, SD = 2.26), followed by Group 3 (Mean = 22.20, SD = 2.53) and Group 1 (Mean = 22.10, SD = 2.38). The median values showed a similar trend in terms of direction, where Group 2 demonstrated the highest median (23.50), whereas both Group 1 and Group 3 showed same median values i.e 22 .00. With respect to variability in the data, the interquartile range (IQR) was highest in Group 3 (IQR = 5.25), which indicated greater dispersion in performance among older participants, whereas Group 2 demonstrated relatively lower variability (IQR = 2.75).

Figure 4.2 *Box-and-Whisker Plot Showing Distribution of Raw Scores Across Groups*



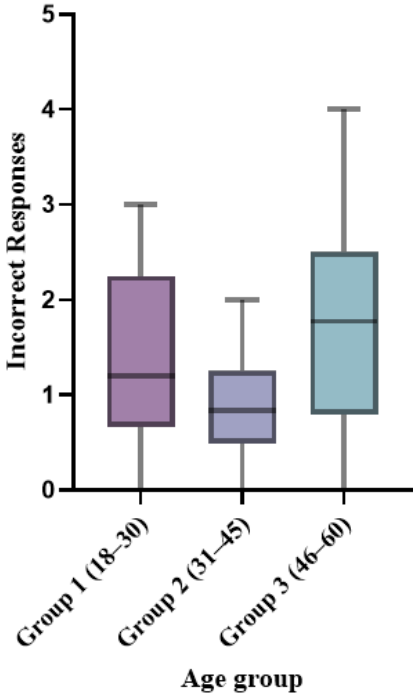
To analyze the Raw Scores (correct + partially correct responses), a similar directional pattern was observed. Group 2 demonstrated the highest mean (Mean = 24.58, SD = 1.76), followed by Group 1 (Mean = 23.78, SD = 2.03), and Group 3 (Mean = 23.65, SD = 2.22). The median values were closely aligning with the mean values, which suggested that there is consistency in central tendency. Group 2 also demonstrated the lowest dispersion (IQR = 2.63), whereas both Group 1 and Group 3 showed slightly higher variability (IQR = 3.25).

Figure 4.3 *Box-and-Whisker Plot Showing Distribution of Partial Responses Across Groups*



With respect to the partial responses, Group 1 demonstrated the highest mean (Mean = 3.35, SD = 1.95), followed by Group 3 (Mean = 2.90, SD = 1.55), while Group 2 demonstrated the lowest mean (Mean = 2.25, SD = 1.37). The median values reflected a similar trend. Notably, Group 2 demonstrated an interquartile range of 0.00, indicating a narrow range of variability of responses, whereas greater variability was observed in Group 1 (IQR = 2.75) and Group 3 (IQR = 2.00).

Figure 4.4 *Box-and-Whisker Plot Showing Distribution of Incorrect Responses Across Groups.*



Accordingly, for studying Incorrect Responses, the highest mean was observed in Group 3 (Mean = 1.80, SD = 1.06), followed by Group 1 (Mean = 1.40, SD = 0.88), whereas Group 2 demonstrated the lowest mean (Mean = 0.65, SD = 0.67). The median values showed a similar directional pattern, with both Group 1 and Group 3 demonstrating a median of 2.00, while Group 2 showed a median of 1.00. Variability, as reflected by the IQR, was highest in Group 3 (IQR = 1.75) and lowest in Group 1 and Group 2 (IQR = 1.00).

In order to verify if there is significant difference across the age groups, appropriate statistical tests were applied. One-way ANOVA was done for Correct responses and Raw scores, since they followed normal distribution, it revealed no statistically significant difference across age groups $F(2,57) = 1.979$, $p > 0.05$.

Similarly, for raw scores, the ANOVA did not show a significant difference, $F(2,57) = 1.243$, $p > 0.05$.

The variables that did not satisfy the assumption of normality, non-parametric analysis was carried out. Hence, the Kruskal-Wallis H test was applied to examine differences across the three age groups for Partial and Incorrect responses.

To check the partial responses, the Kruskal-Wallis test revealed no statistically significant difference across the three age groups, $\chi^2(2) = 4.457$, $p > 0.05$. Although the mean ranks showed that participants in Group 1 (Mean Rank = 34.83) and Group 3 (Mean Rank = 32.45) demonstrated relatively higher partial responses compared to Group 2 (Mean Rank = 24.23), these differences were found to be not statistically significant.

In contrast to this, for incorrect responses, the Kruskal-Wallis test revealed a statistically significant difference across age groups, $\chi^2(2) = 13.824$, $p < 0.05$. The mean rank comparison showed that participants in Group 3 (Mean Rank = 38.60) demonstrates the highest number of incorrect responses, followed by Group 1 (Mean Rank = 33.33), whereas Group 2 showed the lowest mean rank (Mean Rank = 19.58). This indicated that incorrect responses varied significantly across age groups.

To further explore the significant finding for incorrect responses, pairwise comparison was carried using the Mann-Whitney U test.

Between Group 1 and Group 2, a statistically significant difference was observed for incorrect responses ($Z = 2.729$, $p < 0.05$), which indicated that Group 1 elicited significantly more incorrect responses than Group 2. However, partial

responses between these two groups was not of any statistical significance ($Z = 1.907$, $p > 0.05$).

Between Group 1 and Group 3, no statistically significant differences were observed for either partial responses ($U = 180.50$, $Z = 0.543$, $p > 0.05$) or incorrect responses ($U = 161.00$, $Z = 1.113$, $p > 0.05$), indicating that there is identical performance between these two groups.

Between Group 2 and Group 3, a statistically significant difference was observed for incorrect responses ($Z = 3.479$, $p < 0.05$), which indicated that participants in Group 3 produced significantly more incorrect responses than those in Group 2. However, no statistically significant difference was observed for partial responses between these two groups ($Z = 1.698$, $p > 0.05$).

Overall, the findings suggested that even when the number of correct responses and raw scores did not differ significantly across age groups, incorrect responses showed significant age-related variation. Group 2 demonstrated comparatively lesser incorrect responses, whereas Group 3 showed the highest number of incorrect responses. Partial responses also varied in descriptive statistical values, but did not show statistically significant differences across age groups.

4.3 Qualitative analysis

A qualitative analysis was carried out to determine the complexity gradient of the Hindi Auditory Naming Test items. Each response was scored as 1 for correct and 0 for incorrect. The aim of carrying out this analysis was to identify which items received the maximum and minimum correct responses and to check how performance

varied across concrete and abstract stimuli. Items 1 to 17 were concrete stimuli, and the remaining items were abstract stimuli.

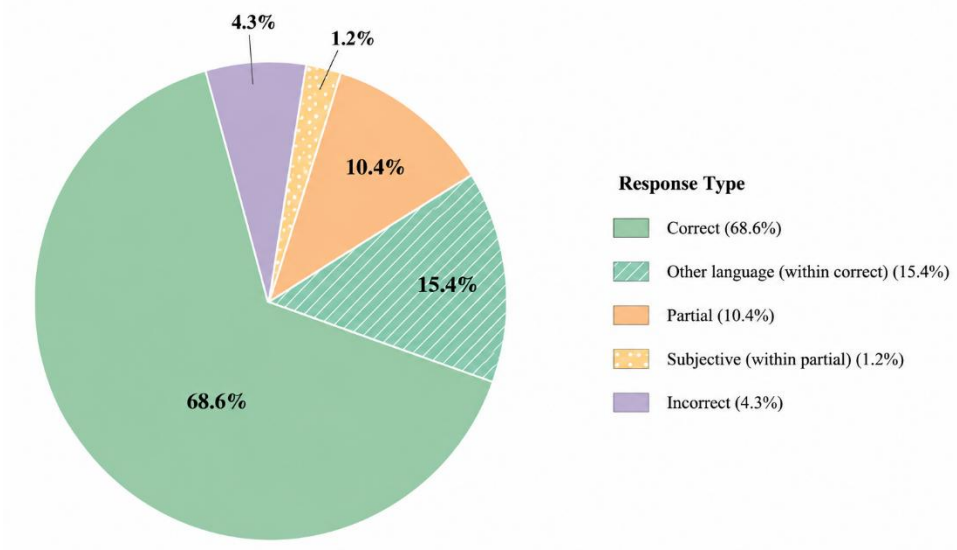
The analysis showed variation in the number of correct responses across items. Item 1, Stim 2, 3, 4, 5, 6, 7, 8, 9, 10, 13, 14, 15, 16, and 17 received the highest number of correct responses, indicating that they were relatively easier for the participants. In contrast, Item 15, 22, 24, 28 have received the lowest number of correct responses, suggesting that it was more difficult. When comparing stimulus types, concrete items showed higher overall accuracy compared to abstract items. A greater number of incorrect responses were observed for abstract stimuli.

This pattern suggests that abstract items required greater lexical retrieval effort than concrete items. The findings indicate that performance decreased as the conceptual complexity of the stimuli increased.

4.3.1 Typological responses

Fig 4.5

Distribution of types of responses in Group I in the study



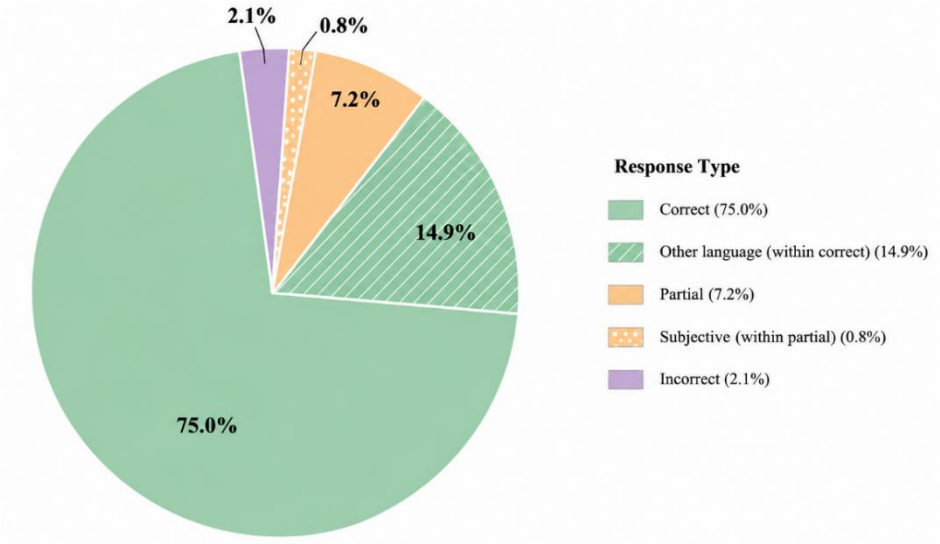
The typological distribution of the response types was analysed in order to assess the qualitative nature of these responses to assess variations across the age groups. To rep llife scenario, responses were classified in different divisions here, such as, correct, partially correct and incorrect categories. This step was done to enhance the ecological validity of the test. Responses that were produced in other language were also considered to be kept under the area of correct responses. On a similar note the subjective responses category was kept inside the area of partially correct responses, which is valid since these responses also reflect partial access of the concept and less precision.

In group 1, the largest portion was seen to be occupied by correct responses (68.6%). This reflects the very nature of neurotypical ability to efficiently retrieve the lexical information. The responses in other language groups within the same category also occupied a significant area (15.4%) in the overall sector graph; this may be attributed to an increase in language mixing. It still be able to suggest that the conceptual knowledge remains intact in these individuals.

Other portions of the sector graph involved, partially correct (10.4%), under which subjective response also covered 1.2% of the area. This reflects that responses to these stimuli had mild lexical difficulties, where participants were able to access semantic features but were not able to produce precise lexical items verbally. Apart from this, incorrect responses covered 4.3% of the total area, which was minimal and reflected a lower frequency of complete failure of retrieval. It is to be noted that these types of responses were mainly seen in abstract response eliciting stimuli.

Fig 4.6

Distribution of types of responses in Group II of the study



Similar trends for response categorization were followed for the second group, responses were categorized into correct, partially correct, and incorrect types. Responses provided in other languages were covered under the category of correct response, which indicates accurate lexical retrieval ability. Similarly, subjective responses were adjusted under partially correct responses, as they reflect reduced lexical precision. These subdomains are visually represented in the given sector graph, within their respective primary categories to maintain both conceptual clarity and analytical consistency.

The distributional area of correct responses in group 2 covered the largest portion of the overall response distribution, which accounts for approximately 75.0% of the total distribution. This distribution shows that participants demonstrated efficient lexical retrieval by being able to access and produce the target response accurately. Responses in other languages contributed to an additional 14.9% of this area, suggesting that participants' bilingual proficiency may affect the response type,

while still being able to retrieve the correct concept, indicating intact semantic knowledge.

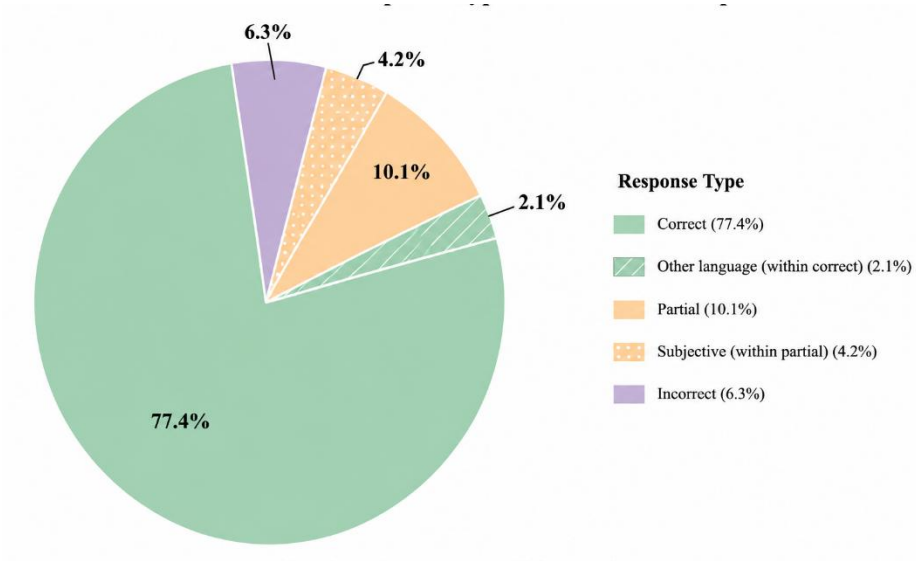
Partially correct responses cover around 7.2 % of the total response area. This indicates that there were limited instances where incomplete or approximate responses were elicited by the participants, where they were able to access some level of target meaning but were unable to produce the exact term. Within this response category, subjective responses covered very small area of approximately 0.8%, which means their responses involved occasional descriptive strategies to answer instead of precise answers.

2.1 % of the total area was occupied by Incorrect responses, that indicated there were fewer instances of complete failure in lexical retrieval overall in the group.

Overall, the response pattern observed in group 2 shows high accuracy and relatively stable performance of the participants of this group. Most area of the sector graph is covered by correct responses and the least by incorrect responses.

Fig 4.7

Distribution of types of responses in Group III of the study



The pattern of responses in group 3 shows a slightly deviant distribution of responses compared to response types of previous age groups. Here also the responses were categorized in same manner, i.e., correct, partially correct and incorrect types, where responses in other language (English) were also included inside the area covered by correct responses, as they reflect accurate conceptual retrieval despite being said in other language. Similarly, subjective responses were considered within the area covered by partial responses as they indicate partial understanding and lesser precision in word retrieval.

In Group 3, correct responses occupied the largest area which was seen in previous groups as well, this shows that participants were generally able to access and retrieve the required lexical item. Responses obtained in other language within this category further supported that the underlying semantic knowledge remained intact, even when the output was not in the expected language.

It is observable that there is a distinction in partially correct responses, which is higher compared to the earlier groups. This shows that participants produced incomplete responses more frequently, which were in approximation to the actual response. This suggests that there was some difficulty in retrieving the exact lexical item for a particular stimulus. Although the general meaning was often accessible, the precision of expression was reduced. The inclusion of subjective responses within this category further highlights the tendency to rely on descriptive or generalized responses when exact word retrieval was not achieved.

In addition to this, incorrect responses occupied a comparatively larger proportion than what was observed in the previous group, this indicated an increase in instances of complete failure of retrieval. This reflects more variability in the performance and this also suggests that participants in this group have experienced more challenges in accessing the correct lexical items.

Taken together, the response pattern in Group 3 reflects a decline in accuracy and an increment in variability. While the ability to understand and access concepts appears to remain adequate, reduction in the efficiency of lexical retrieval can be clearly seen, as seen in the higher proportion of partially correct and incorrect responses.

4.4 Objective 3: To determine inter-rater and test-retest reliability by analysing 10% of the collected data.

4.4.1 Inter-rater reliability

Inter-rater reliability was assessed to check the consistency of scoring across different raters. For this purpose, 10% of the total dataset was selected, which included responses from six participants out of the total sample of sixty participants. These responses were individually evaluated by three raters. The ratings were given using a three-point scoring system, where responses were categorized as incorrect (0), partially correct (0.5), and correct (1).

To test the level of agreement among the raters, Cronbach's alpha was calculated using the reliability analysis procedure. The other alternative for this test was Kappa's coefficient, which was not recommended since the total number of raters was 3, hence the investigator decided to assess the data using Cronbach's alpha. In addition, Intraclass Correlation Coefficient (ICC) was also computed to further assess the degree of agreement between raters.

The analysis was conducted on 6 valid cases. Cronbach's alpha value of 0.991 indicated excellent reliability (Cronbach's alpha values above 0.70 are generally considered acceptable, values above 0.80 indicate good reliability, and values above 0.90 indicate excellent reliability).

Further support for this finding was seen in the Interclass correlation (ICC) result values. ICC values of 0.991 showed excellent agreement (ICC values above 0.75 are considered good, and values above 0.90 indicate excellent agreement). The

single-measures ICC was 0.972, with a 95% confidence interval ranging from 0.965 to 0.978, this indicated very high agreement at the individual rater level as well.

The crosstab analysis also demonstrated strong agreement between raters. Most of the responses were consistently rated as correct (score of 1) for the correct response category, across the raters, with minimum variations observed in partially correct and incorrect categories.

Table 4.6

Test–Retest Reliability Analysis of HANT Scores

Participant No.	First Administration Score	Second Administration Score	Difference Score
P1	27	28	1
P2	29	29	0
P3	25	26	1
P4	30	30	0
P5	28	27	1

Participant No.	First Administration Score	Second Administration Score	Difference Score
P6	26	26	0

Table 4.6a

Summary of Test–Retest Reliability Statistics for HANT

Reliability Measure	Value
Test–Retest Reliability Coefficient (r)	0.96
Significance Level (p-value)	< 0.001

Table 4.5 and Table 4.5a, presents the test–retest reliability analysis of the Hindi Auditory Naming Test (HANT) conducted on 10% of the collected sample. The table compares participant scores obtained during the first and second administrations of the test and depicts the consistency of performance across repeated testing sessions. The obtained reliability coefficient demonstrated excellent temporal stability, indicating that the HANT yields consistent results over time and can therefore be considered a reliable measure for assessing auditory lexical retrieval abilities.

Overall, the results for inter rater reliability testing showed that the scoring scheme used in the study is highly reliable, and also have a strong agreement across raters. The high values of

Cronbach's alpha and ICC indicated that the ratings are consistent and suitable for further analysis.

4.4.2 Test-Retest Reliability

Test–retest reliability was assessed to examine the stability and consistency of the test over time. This method helps to determine whether the test produces similar results when it is administered on the same participants over two different timelines. For this purpose, responses from six participants were analyzed by comparing their performance at Time 1 (T1) that was the first time when the data was acquired from each participant and Time 2 (T2) when the participants were subjected for reliability testing. The time duration between T1 and T2 was 5 months, suggesting that practice effect was minimal or not there.

To measure the level of agreement between the two testing sessions, Cohen's Kappa was used. This statistical test is appropriate to be used for this objective because it evaluates the degree of agreement beyond the chance for categorical or discrete data. Kappa values generally range from 0 to 1, where values below 0.40 indicate poor agreement, values between 0.41–0.60 indicate moderate agreement, values between 0.61–0.80 indicate good agreement, and values above 0.80 indicate excellent agreement.

The obtained value of Kappa was 0.419 for correct responses, which indicates moderate agreement between the two testing sessions . This shows that participants showed a reasonable level of consistency in producing correct responses across time, even though some variabilities were present.

The Kappa value obtained for partially correct was 0.217, indicating fair to low agreement . This suggests that partially correct responses were less stable across testing sessions, this is possible since these responses involve approximation. According to McCauley, Rebecca J. & Swisher, Linda (1984), non binary responses ((e.g., partial correctness) shows more variability and reduced reliability compared to clearer binary scored categories (eg. correct/incorrect).

For incorrect responses, the Kappa value was 1.000, indicated perfect agreement . This shows that incorrect responses were highly stable across both testing sessions, with participants consistently failing on the same items.

Furthermore, the raw scores Kappa value was again 0.419, reflecting moderate level of agreement. This indicates that overall performance across 2 timeline was relatively consistent, even though minor variations were observed.

Table 4.7

Inter-Rater Reliability Analysis of HANT Scores

Participant No.	**Rater 1 Score**	**Rater 2 Score**	**Rater 3 Score**
P1	28	28	27
P2	29	29	29
P3	26	25	26

Participant No.	Rater 1 Score	Rater 2 Score	Rater 3 Score
P4	30	30	30
P5	27	28	27
P6	26	26	26

Table 4.7a

Summary of Inter-Rater Reliability Statistics for HANT

Reliability Measure	Value
Inter-Rater Reliability Coefficient (ICC)	0.98
Significance Level (p-value)	< 0.001

Table 4.6 and Table 4.6a, presents the inter-rater reliability analysis of the Hindi Auditory Naming Test (HANT). The table compares the scores assigned independently by three raters for the same set of participant responses. The high inter-rater reliability coefficient indicates strong agreement among raters, demonstrating that the scoring procedure of the HANT is objective, consistent, and minimally influenced by individual examiner variability. These findings support the reliability and scoring consistency of the developed test.

Overall, the results suggested that the test showcase moderate to high test–retest reliability. While correct responses and overall performance showed appropriate stability, and on the other hand partially correct responses were more variable, which is expected due to their less precise nature. The near perfect agreement observed for incorrect responses further supports the consistency of certain response patterns across the 2 timelines.

Chapter 5

Discussion

The present study was conducted with the objectives of developing the content of an Auditory Naming Test (HANT) and validating it to ensure linguistic appropriateness and cultural relevance. Conduction of field testing on a controlled sample was also done to examine performance variations across three age groups (18–30, 31–45, and 46–60 years), and also determining the inter-rater and test–retest reliability of the test by analysing 10% of the collected data .

The development and validation of the HANT taps onto an important gap in existing assessment tools, as most conventional naming tests rely on visual stimuli and may not be able to adequately capture auditory lexical retrieval, which is more representative of everyday communication (Hamberger & Seidel, 2003; Fergadiotis & Wright, 2011). This distinction is also supported by modality-specific processing frameworks, which says that auditory and visual naming tasks engage partially overlapping but functionally distinct neural and cognitive pathways (Indefrey & Levelt, 2004; Tomaszewski Farias et al., 2005). In continuation, the present study specifically addressed this limitation by systematically designing and validating the content of the Hindi Auditory Naming Test (HANT) to ensure that the stimuli are linguistically accurate, culturally meaningful and functionally relevant for the target population .

The first step was to design the stimuli of the test, which was carried out in a structured manner to ensure that each question effectively elicit a specific lexical target response (using a single word as a response). Initially a set of 31 auditory

stimulus questions was designed in Hindi. The content validation process played a crucial role in establishing this relevance. By involving expert speech-language pathologists and evaluating each item across the 5 parameters such as clarity, relevance, ambiguity, familiarity, and appropriateness. Both quantitative and qualitative feedback were acquired from the evaluators. Thus, the study ensured that the auditory prompts were capable of eliciting the intended lexical responses effectively. The use of the Item-wise Content Validity Index (I-CVI) and Scale- level Content Validity Index (S-CVI) as a quantitative measure, further gave strength to this process by providing an objective measure of agreement among experts, which allowed only those items with adequate consensus to be retained while others were revised or excluded (Lynn, 1986).

The results of the content validation demonstrated that experts have a high level of agreement for most of the items. Most of the stimulus questions (stimulus 1 to 30) obtained an I-CVI value of 0.78 or above, which indicated acceptable content validity for those questions (except stimulus 31). A large number of the items demonstrated strong agreement (I-CVI $\geq$ 0.78-1.00), suggesting that they were clear, relevant, and appropriate for eliciting the intended responses. The overall Scale-level Content Validity Index (S-CVI) and universal agreement was also found to be high, that indicates good content validity at the test level.

However, as mentioned previously few items (stimulus number 16, 17 and 31) received comparatively lower I-CVI values because of ambiguity in wording, presence of multiple possible responses, or reduced familiarity. These items were carefully reviewed based on expert's qualitative feedback. Modifications were made by simplifying sentence structure, increasing semantic specificity, and replacing less

familiar or culturally less relevant words. No items were removed directly without the review of the experts, these items were considered to be revised wherever possible since they are functionally valuable.

Once this process was over, where the stimuli were refined and modified, one item i.e stimulus number 31, which did not receive acceptance from the experts after revision was excluded. In the end a final set of 30 stimulus questions was established. all items in this set had acceptable I-CVI and S-CVI thresholds, which means they cleared on all the required parameters, which were required for a stimulus to be retained.

In conclusion, the results of content validation for the Hindi Auditory Naming Test (HANT) indicate good content validity, with most items showing strong agreement among experts. The I-CVI and S-CVI helped to provide a systematic and objective review of the test items. The revision of a few items, based on the experts' observations, made them more relevant, clearer, and culturally suitable for the assessment. The Final Set of 30 HANT questions was found to be effective and linguistically appropriate for eliciting the required responses from older adults. The HANT findings confirm that the stimuli used in the test are suitable for assessing auditory naming skills, providing a solid foundation for assessment and for further testing and validation.

This in depth and structured validation process is a necessary part for development of any test material, especially for an auditory naming test because, here listener has to listen to the stimuli and then decode the auditory input, which is then followed by maintaining it in working memory. After this process, the semantic meaning is extractable and the retrieval of an appropriate lexical item can be done to

respond immediately (Gainotti, 2014; Nickels, 2002). If there is any ambiguity in the stimulus, the processing demand may increase, inturn the cognitive taxing may also increase, which hinders the assessment of true lexical retrieval ability. In this study, the author kept the stimuli concise, clear, and semantically focused, which may help in minimizing these extraneous variables (cognitive taxing), and focus only on the auditory lexical access.

Along with these considerations, controlled psycholinguistic variables (lexical frequency, semantic category, and concreteness) also gave a strong foundation and balanced graded structure to the test, resulting is higher sensitivity of the tool. This also allowed us to assess variations in lexical retrieval across different difficulty levels, which is important for clinical assessment and a research point of view.

The second objective of this study was to examine variations in performance across three age groups (18-30 years, 31-45 years, and 46-60 years) using the Hindi Auditory Naming Test (HANT), as lexical retrieval related cognitive-linguistic functions change as a function of age.

The performance was analysed based on number of correct, partially correct, and incorrect responses, obtained from each individual, along with their raw scores.

The findings showed that there was no statistically significant difference in the number of correct responses and raw scores across the three age groups (Correct responses: $F(2,57) = 1.979$, $p > 0.05$; Raw scores: $F(2,57) = 1.243$, $p > 0.05$). This suggests that overall accuracy in auditory naming remains relatively stable across adulthood. In other words, the ability to retrieve the correct lexical item, appears to be preserved in neurotypical individuals across the selected age range.

However, when the different response types were analysed in detail, a more meaningful pattern emerged. Incorrect responses showed a statistically significant difference across age groups ($\chi^2(2) = 13.824$, $p < 0.05$), whereas partially correct responses did not show statistically significant variation ($\chi^2(2) = 4.457$, $p > 0.05$). Specifically, the middle age group (31-45 years) demonstrated the lowest number of incorrect responses (Mean Rank=19.58), while the older age group (46–60 years) showed the highest number of incorrect responses (Mean Rank = 38.60). The younger group (18-30 years) showed performance that was almost intermediate (Mean Rank = 33.33).

The observed pattern suggests that, even though overall accuracy may be similar across age groups, there are slight differences in lexical retrieval efficiency among subjects. The older age group showed increased incorrect responses, which indicates the reduced ability to access the appropriate lexical items but not the meaning of the items. These results are consistent with the earlier research based on cognitive ageing, which shows that as age advances the semantic knowledge is preserved but the word retrieval ability declines (Deborah M. Burke et al., 1991; Meredith A. Shafto et al., 2007).

Similar findings have been reported in previous studies examining age-related changes in naming performance. For instance, a study by Meredith A. Shafto et al. (2007), which included participants across a wide adult age range (18–80 years, n = 48), used a picture naming task and reported that older adults showed increased word-finding difficulties despite preserved semantic kno wledge. The authors suggested that these changes are primarily due to reduced efficiency in lexical retrieval rather than degradation of semantic representations. This is comparable to the findings of the

present study, where no significant difference was observed in correct responses, but an increase in incorrect responses was noted in the older age group.

From a theoretical perspective, semantic activation, lexical retrieval, and phonological retrieval are involved in the process of lexical retrieval. The present study indicates that at the level of lexical access or lexical selection, there may be age-related changes, such as a delay in the retrieval process of the required lexical items. However, in the same population, the conceptual knowledge known as the semantic system remains intact. A study by Hamberger and Seidel (2003) revealed that auditory naming places a greater cognitive load on the word retrieval processes with respect to accuracy and efficiency. They found that, compared to visual naming tasks, subtle word finding difficulties are far more sensitive in auditory naming tasks. Though both studies are different with respect to their methodology and linguistic context from the present study, their findings are comparable, as both studies highlight the role of auditory modality in the lexical retrieval process.

With respect to partially correct responses, descriptive differences were observed which showed relatively higher values in the younger and older groups compared to the middle age group, but still these differences were not statistically significant ($Z=1.907$, $p > 0.05$; $Z=0.543$, $p > 0.05$; $Z=1.698$, $p > 0.05$). Partially correct responses included semantic substitutions, phonological approximations, and circumlocutions, these imperfections in the response demonstrate incomplete or imprecise lexical retrieval. The lack of statistical significance here suggests that such responses may also occur across all age groups as a part of normal variability in lexical access.

An important finding of the present study is the inclusion of graded response categories (correct, partially correct, incorrect) rather than simple binary scoring system. This mechanism allowed a more sensitive analysis of lexical retrieval performance in participants. It can be noted that the presence of more incorrect responses highlights the age-related changes, since the error patterns are more evident in older age groups, when compared to the scores that only reflect the accuracy related measure.

In the later qualitative analysis, a similar pattern was observed when the complexity gradient evaluation was carried out, which supported this finding. It was seen that concrete stimuli elicited higher accuracy compared to the abstract ones. Since concrete items have more sensory experience for an individual, they are said to be accessed faster and more easily, whereas abstract words depend more on contextual and associative networks, which increases processing demands. Hence, It was also seen that abstract stimuli received the largest number of partially correct and/or incorrect responses. It can be inferred that the abstract stimuli demand more cognitive taxing, comparatively. Though there are contradicting researches exist in the literature, some studies do explain the concept of concreteness effect, where concrete words are processed more efficiently than abstract words (e.g., Paivio, 1991; Brysbaert et al., 2014). Hence, we can conclude that the type of stimulus plays an important role in naming and its performance, which is independent of the age of the participant.

The qualitative analysis provided more detailed insights while analysing the responses received from the participants across age groups. Across the groups, the largest proportion of responses were towards the correct responses, indicating that older adults had more incorrect and partially correct responses compared to the

younger adults. This indicates that while the conceptual understanding of older adults remains preserved, their precision in responses is compromised by lexical retrieval as they get older.

Another important observation was that, for several questions, some adults responded in another known language; although the responses were in a different language, they were appropriate and considered correct. This discusses the multilingual nature of the Indian population and indicates that, even when code-switching or code-mixing occurs, the deeper structure of the response remains intact. In the Indian context, this can play a major role, as the majority of the population uses more than one language.

Similar patterns are observed in the literature, such as, it is reported that semantic knowledge remains stable across adulthood, on the other hand retrieving efficiency becomes poorer with age (e.g., Salthouse, 2010; Craik & Bialystok, 2006). However, the present study findings suggest no significant difference across age when it comes to correct responses. This disagreement may have arisen due to some notable parameters, such as culturally appropriate stimuli along with controlled sample characteristics, it may also be due to the use of a non-binary scoring system that involved partially correct responses, which may have resulted in the sensitivity of the assessment tool. On the other hand, the trend may not have been visible due to a smaller sample size.

The findings of this study emphasise that throughout adulthood, auditory naming ability remains stable and age-related subtle changes are observed in terms of the various responses and erroneous productions. It also emphasises why it is important to use the graded scoring system while assessing the lexical retrieval ability.

The third objective of the study was to determine the inter-rater reliability and test-retest reliability of the Hindi Auditory Naming Test by subjecting 10% of the collected data for reliability analysis.

The results may indicate that HANT has received excellent inter-rater reliability. This can be inferred from the statistical values of Cronbach's alpha, which was α= 0.991, and also through the interclass correlation coefficient (ICC), which was also 0.991. These values are indicative of a high level of agreement among the raters. It also indicates that the scoring system used for the assessment is consistent and reliable.

The findings are important, particularly because the test uses a three-point scoring system that requires a certain degree of freedom for some subjective judgment for particularly correct responses. This way,, the clinician's subjective judgement can play an important role in the assessment. The responses that meet the strict criteria could be rejected, but now receive a score based on the clinician's proficiency. However, the high agreement among raters (Cronbach's α =0.991; ICC =0.991; single-measures ICC =0.972, 95% CI =0.965-0.978) indicates that the scoring rules were clearly defined and consistently applied by all evaluators. The high inter-rater reliability also strongly supports the test's replicability, meaning it would yield the same results upon re-administration by the same or different clinicians on the same subject, which is the primary necessity for any clinical or research tool..

Test-retest reliability was assessed to determine the stability of the test over time. The findings showed moderate agreement for correct responses and overall raw scores (κ=0.419), whereas partially correct responses showed lower agreement (κ= 0.217), and incorrect responses showed perfect agreement (κ =1.000).

The moderate reliability observed for correct responses suggests that participants were reasonably consistent in their performance across time, although some variability was present. This variability may be affected by factors such as attention, fatigue, or slight fluctuations in retrieval efficiency.

The lower reliability observed for partially correct responses can be explained by their own variable nature. These responses depend on approximation strategies, such as semantic substitutions or circumlocutions, which may differ across testing sessions. Therefore, some degree of variability in partially correct responses is expected. (Nickels, 2002; Dell et al., 1997)

In contrast, incorrect responses showed near perfect agreement across sessions (κ=1.000), which indicated that items which were not retrieved correctly happened to remain consistently difficult for the participants. This suggests that certain lexical items may be more difficult to retrieve and may have led to stable patterns of incorrect responses.

5.1 Proposed Protocol for the Development and Administration of the Hindi Auditory Naming Test (HANT)

5.1.1 Stimulus-Related Guidelines

HANT needs sufficient culturally and linguistically appropriate stimulus items to assess the auditory naming, without causing the participant fatigue. In this study, a total of 30 stimulus items were found suitable for the administration. The assessment stimuli should include both abstract and concrete words to reduce the ceiling effect and balance the linguistic complexity. Additional words from different semantic categories, including both frequent and infrequent lexical items, should be included to

provide a comprehensive assessment of lexical retrieval. Stimulus questions should be short, simple, and easy to understand so that participants can process the auditory information without excessive cognitive load. Preferably, each question should contain not more than 7-9 words. Minimally complex or simple sentences can be used to frame the question, with a preferred mean length of 7-9 words, so that participants can process the auditory information without excessive cognitive load. Standardised instructions should be provided before the test, and a few practice items may be administered to familiarize participants with the task.

The test may be administered through two modes, first may involve live presentation by the examiner, or the second may involve a recorded auditory stimulus, depending on the feasibility. The use of recorded stimuli may help maintain uniformity across participants and testing settings.

5.1.2 Administration-Related Guidelines

The estimated administration time for HANT is approximately 15–25 minutes, this may depend and vary according to the age and processing speed of the participant. Each stimulus may be considered to be presented with an interval of about 5 seconds, and participants can be given up to 10 seconds to respond (in case of neurotypical)

Participants should be instructed to respond as quickly and accurately as possible. No corrective feedback should be provided during testing to avoid response bias. The testing environment should be quiet and free from distractions to ensure optimal auditory processing and concentration.

If a participant is unable to respond, a structured cueing hierarchy may be used. Semantic cues may be provided first, followed by phonemic cues and finally forced-

choice options if required. Responses following cues may provide additional clinical information regarding the nature and severity of lexical retrieval difficulties.

Examiners administering the test should be familiar with standardized administration procedures to ensure consistency across testing sessions.

5.1.3 Scoring-Related Guidelines

Participant responses should be systematically recorded and categorized under the following categories:

- Correct response
- Correct response after cue
- Partially correct response
- Incorrect response
- No response

Cue-based scoring may help identify whether the difficulty primarily involves lexical retrieval or broader language-processing deficits. Recording response latency may also provide additional information regarding naming efficiency and processing speed.

Future work can focus on establishing normative data, confirming reliability, and confirming that HANT works well across different clinical groups. This would help make it more useful in clinical practice. It may serve as a helpful clinical tool for assessment of auditory naming difficulties in people with dementia, aphasia, traumatic brain injury, and other neurogenic communication disorders. During the test, certain changes may be required depending on the severity of the conditions. Extra response time, stimulus repetition, number of trials, and cueing support may be provided when required. To minimise fatigue and maintain attention, the number

of items may be reduced when required. Further research can examine adapting HANT to other Indian languages to assess lexical retrieval abilities across language groups. During adaptation, culturally and linguistically appropriate words can be included to increase familiarity.

While carrying out the adaptation of this test in other Indian languages, the WHO's translation guidelines are advised to be followed. This may involve forward translation, expert review, back translation, pre-testing, and cultural adaptation to make sure that the test remains meaningful and relevant for the regions across the country and languages as well. This process will ensure that the test remain suitable for a multilingual population, especially in the Indian context, which will further be helpful in clinical and research settings.

Chapter 6

Summary and Conclusion

The present study aimed to develop and validate the Hindi Auditory Naming Test (HANT) to measure the lexical revival abilities in the neurotypical adults through auditory mode. Traditional naming tests usually rely primarily on the visual confrontation tasks. These tasks may not completely reflect how people actually communicate in real life. The present study aimed to design a tool which is culturally and linguistically appropriate and focuses on the naturalistic language processing in the Hindi speaking population.

The development of the Hindi auditory naming test involved a multistep process, which involved three phases. In the first phase, a set of 31 stimulus questions was made in the Hindi language to elicit a single word response for each. These stimuli were curated to replicate everyday communicational needs, which involve auditory verbal modality. These stimuli included a variety of lexical semantic aspects, such as abstract vs concrete items and frequent vs infrequent items. The second phase of the study was to validate the content; the stimuli were sent to five speech-language pathologists to validate the content on several parameters, after which the I-CVI and S-CVI were calculated, which are the standard measures to calculate the content validity index. Only those items which met the acceptable threshold were retained or modified based on the feedback. The last and final phase was to administer this test on 60 neurotypicals and to assess if the naming abilities change as a function of age.

The responses of the participants were analysed mainly to check how accurately participants named the items, finding or sorting correct or incorrect

responses. To understand the patterns of lexical retrieval ability qualitative error analysis was performed. Statistical analysis was conducted to assess If the age affected the naming performance. Reliability was measured for both inter-rater and test-retest reliability on a subset of the collected data

The final outcome of this study shows that the Hindi Auditory Naming test (HANT) has good content validity, which suggests that the stimuli hold good linguistic and culturally relevant stimuli that are relevant for Hindi- speaking adults. In the reliability analysis, a high level of agreement across raters and across timelines was observed, which reflects that the test is stable and consistent with the scoring procedure.

The results indicated that auditory naming was significantly affected by individuals' ages. The older adults showed comparatively poorer naming accuracy than younger adults. The results were consistent with earlier psycholinguistic and neurocognitive research findings, which show that lexical retrieval ability declines with advancing age, due to changes in processing speed and in how semantic network activation occurs in older adults.

In a nutshell, the study successfully established the Hindi Auditory Naming test as a valid and reliable tool for assessing auditory lexical retrieval ability in the Hindi speaking population. This test can be used to assess modality specific naming performance and can also be used in conjunction with other language tests to provide insights into real-life language processing demands.

Implications of the Study

- The HANT serves as a clinically relevant tool to assess auditory naming, and more precisely reflects the day to day communication demands and needs than the visual naming tasks..
- This tool can be used to identify and differentially diagnose between the language disorders such as dementia, aphasia, and mild cognitive impairment. The auditory processing and lexical retrieval are differentially affected in these conditions.
- This study emphasises that lexical retrieval varies with age, which is an important consideration in the assessment and the interpretation of the naming performance.
- The clinical utility is enhanced by incorporating qualitative analysis, which provides insights into the underlying processing deficits rather than relying only on score accuracy.

Limitations of the Study

- The sample size taken was relatively small, i.e., a total of 60 participants were taken for the study, which may affect the generalizability of the test.
- Educational levels of the participants were not kept strict across the age groups, particularly in older adults, which may have influenced performance.
- The study focused only on accuracy measures, other measures, like reaction time and tip of the tongue phenomenon, could provide additional sensitivity to lexical retrieval processes.

Future Directions

- Future studies may involve reaction time to provide a more detailed understanding of lexical access dynamics across different age groups.
- The test can be validated on clinical populations such as individuals with aphasia, dementia, and traumatic brain injury.
- Comparative studies between auditory and visual naming tasks can further explore modality-specific differences in lexical retrieval.
- Expansion of HANT into other Indian languages can contribute to cross-linguistic standardization of auditory naming tools.

References

Abreu-Mendoza, R. A., Arias-Trejo, N., & López-Ornat, S. (2020). Naming abilities and lexical access across the lifespan: Evidence from auditory and visual naming tasks. Journal of Psycholinguistic Research, 49(5), 793–810. https://doi.org/10.1007/s10936-020-09710-5

Annamalai, E. (2001). *Managing multilingualism in India: Political and linguistic manifestations*. Sage Publications.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baldo, J. V., Schwartz, S., Wilkins, D., & Dronkers, N. F. (2006). Role of frontal *versus* temporal cortex in verbal fluency as revealed by voxel-based lesion symptom mapping. *Journal of the International Neuropsychological Society*, *12*(6), 896–900. https://doi.org/10.1017/S1355617706061078

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart Pictures: Effects of Age of Acquisition, Frequency, and Name Agreement. *The Quarterly Journal of Experimental Psychology A*, *50*(3), 560–585. https://doi.org/10.1080/027249897392026

Bates, E., Devescovi, A., & Wulfeck, B. (2001). Psycholinguistics: A Cross-Language Perspective. *Annual Review of Psychology*, *52*(1), 369–396. https://doi.org/10.1146/annurev.psych.52.1.369

Benson, D. F., & Ardila, A. (1996). *Aphasia: A Clinical Perspective*. Oxford University PressNew York, NY. https://doi.org/10.1093/oso/9780195089349.001.0001

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, Aging, and Cognitive Control: Evidence From the Simon Task. *Psychology and Aging*, *19*(2), 290–303. https://doi.org/10.1037/0882-7974.19.2.290

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. Journal of Cognitive Neuroscience, 17(6), 905–917. https://doi.org/10.1162/0898929054021102

Bürki, A., & Madec, S. (2022). Picture-word interference in language production studies: Exploring the roles of attention and processing times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(7), 1019–1046. https://doi.org/10.1037/xlm0001098

Bürki, A., & Madec, S. (2023). *Picture-word interference in language production studies: Exploring the roles of attention and processing times* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2303.09201

Caramazza, A., & Hillis, A. E. (1990). Spatial representation of words in the brain implied by studies of a unilateral neglect patient. *Nature*, *346*(6281), 267–269. https://doi.org/10.1038/346267a0

Caramazza, A., & Shelton, J. R. (1998). Domain-Specific Knowledge Systems in the Brain: The Animate-Inanimate Distinction. *Journal of Cognitive Neuroscience*, *10*(1), 1–34. https://doi.org/10.1162/089892998563752

Damian, M. F., & Martin, R. C. (2001). Semantic and phonological codes interact in single word production. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(2), 345–361. https://doi.org/10.1037/0278-7393.27.2.345

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801–838. https://doi.org/10.1037/0033-295X.104.4.801

Farias, S. T., Mungas, D., & Jagust, W. (2005a). Degree of discrepancy between self and other-reported everyday functioning by cognitive status: Dementia, mild cognitive impairment, and healthy elders. *International Journal of Geriatric Psychiatry*, *20*(9), 827–834. https://doi.org/10.1002/gps.1367

Farias, S. T., Mungas, D., & Jagust, W. (2005b). Degree of discrepancy between self and other-reported everyday functioning by cognitive status: Dementia, mild cognitive impairment, and healthy elders. *International Journal of Geriatric Psychiatry*, *20*(9), 827–834. https://doi.org/10.1002/gps.1367

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. Aphasiology, 25(11), 1414–1430. https://doi.org/10.1080/02687038.2011.603898

Gainotti, G. (2014). Why are the right and left hemisphere conceptual representations different? Behavioural Neurology, 2014, Article 603134. https://doi.org/10.1155/2014/603134

Georgiou, G. K., Parrila, R., Cui, Y., & Papadopoulos, T. C. (2013). Why is rapid automatized naming related to reading? Journal of Experimental Child Psychology, 115(1), 218–225. https://doi.org/10.1016/j.jecp.2012.10.015

Goodglass, H., & Wingfield, A. (Eds.). (1997). Anomia: Neuroanatomical and cognitive correlates. Academic Press.

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., Ogar, J. M., Rohrer, J. D., Black, S., Boeve, B. F., Manes, F., Dronkers, N. F., Vandenberghe, R., Rascovsky, K., Patterson, K., Miller, B. L., Knopman, D. S., Hodges, J. R., Mesulam, M. M., & Grossman, M. (2011). Classification of primary progressive aphasia and its variants. *Neurology*, *76*(11), 1006–1014. https://doi.org/10.1212/WNL.0b013e31821103e6

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395. https://doi.org/10.1037/0033-295X.101.3.371

Hamberger, M. J., Goodman, R. R., Perrine, K., & Tamny, T. (2001). Anatomic dissociation of auditory and visual naming in the lateral temporal cortex. *Neurology*, *56*(1), 56–61. https://doi.org/10.1212/WNL.56.1.56

Hamberger, M. J., McClelland, S., McKhann, G. M., Williams, A. C., & Goodman, R. R. (2007). Distribution of Auditory and Visual Naming Sites in Nonlesional Temporal Lobe Epilepsy Patients and Patients with Space-Occupying Temporal Lobe Lesions. *Epilepsia*, *48*(3), 531–538. https://doi.org/10.1111/j.1528-1167.2006.00955.x

Hamberger, M. J., & Seidel, W. T. (2003). Auditory and visual naming tests: Normative and patient data for accuracy, response time, and tip-of-the-tongue. *Journal of the International Neuropsychological Society*, *9*(3), 479–489. https://doi.org/10.1017/S135561770393013X

Henry, J. D., & Crawford, J. R. (2004). A Meta-Analytic Review of Verbal Fluency Performance Following Focal Cortical Lesions. *Neuropsychology*, *18*(2), 284–295. https://doi.org/10.1037/0894-4105.18.2.284

Henry, J. D., Von Hippel, W., Molenberghs, P., Lee, T., & Sachdev, P. S. (2016). Clinical assessment of social cognitive function in neurological disorders. *Nature Reviews Neurology*, *12*(1), 28–39. https://doi.org/10.1038/nrneurol.2015.229

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1–2), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Kempen, G. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, *14*(2), 185–209. https://doi.org/10.1016/0010-0277(83)90029-X

Kertesz, A. (2006). Western Aphasia Battery–Revised (WAB-R). Pearson Assessment.

Khwaileh, F. A., Mustafawi, E., Herbert, R., & Howard, D. (2018). Auditory naming performance in Arabic-speaking adults: Effects of age, education, and lexical variables. Clinical Linguistics & Phonetics, 32(10), 901–918. https://doi.org/10.1080/02699206.2018.1441394

Kohnert, K. (2004). Cognitive and cognate-based treatments for bilingual aphasia: A case study. *Brain and Language*, *91*(3), 294–302. https://doi.org/10.1016/j.bandl.2004.04.001

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*(6), 223–232. https://doi.org/10.1016/S1364-6613(99)01319-4

Lynn, M. R. (1986). Determination and Quantification Of Content Validity: *Nursing Research*, *35*(6), 382???386. https://doi.org/10.1097/00006199-198611000-00017

Malow, B. A., Lin, X., Kushwaha, R., & Aldrich, M. S. (1998). Interictal Spiking Increases with Sleep Depth in Temporal Lobe Epilepsy. *Epilepsia*, *39*(12), 1309–1316. https://doi.org/10.1111/j.1528-1157.1998.tb01329.x

Meyer, A. S. (1996). Lexical Access in Phrase and Sentence Production: Results from Picture–Word Interference Experiments. *Journal of Memory and Language*, *35*(4), 477–496. https://doi.org/10.1006/jmla.1996.0026

Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology*, *16*(10–11), 935–979. https://doi.org/10.1080/02687030244000563

Randolph, C., Lansing, A. E., Ivnik, R. J., Cullum, C. M., & Hermann, B. P. (1999). Determinants of Confrontation Naming Performance. *Archives of Clinical Neuropsychology*, *14*(6), 489–496. https://doi.org/10.1093/arclin/14.6.489

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*(6), 1090–1098. https://doi.org/10.3758/BF03196742

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1–3), 107–142. https://doi.org/10.1016/0010-0277(92)90041-F

Saffran, N. M. E. M. (1997). Language and Auditory-verbal Short-term Memory Impairments: Evidence for Common Underlying Processes. *Cognitive Neuropsychology*, *14*(5), 641–682. https://doi.org/10.1080/026432997381402

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00772

Strijkers, K., & Costa, A. (2011). Riding the Lexical Speedway: A Critical Review on the Time Course of Lexical Selection in Speech Production. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00356

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., Lu, C. C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., … Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, *51*(2), 247–250. https://doi.org/10.1016/j.jml.2004.03.002

Tomaszewski Farias, S., Mungas, D., & Jagust, W. (2005). Degree of discrepancy between self- and other-reported everyday functioning by cognitive status: Dementia, mild cognitive impairment, and healthy elders. International Journal of Geriatric Psychiatry, 20(9), 827–834. https://doi.org/10.1002/gps.1367

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. Brain, 107(3), 829–854. https://doi.org/10.1093/brain/107.3.829

Wetherby, A. M., & Prizant, B. M. (1993). Communication and Symbolic Behavior Scales. Paul H. Brookes Publishing.

Wingfield, A., & Goodglass, H. (1990). Propositional speech and language processing in aphasia. In Y. Joanette & H. Brownell (Eds.), *Discourse ability and brain damage: Theoretical and empirical perspectives* (pp. 139–152). Springer.

APPENDIX I

All India Institute of Speech and Hearing
(An autonomous Institute under the
Ministry of Health and Family Welfare, Govt. of India)
Center of Excellence - Assessed & Accredited by NAAC with 'A' Grade
ISO 9001:2015 Implemented Institute
Manasagangothri, Mysuru-570006

ಅಖಿಲ ಭಾರತ ವಾಕ್ ಮತ್ತು ಶ್ರವಣ ಸಂಸ್ಥೆ
ಮಾನಸಗಂಗೋತ್ರಿ, ಮೈಸೂರು-570006

अखिल भारतीय वाक् श्रवण संस्थान
मानसगंगोत्री, मैसूरु - 570 006

INSTITUTE REVIEW BOARD (IRB) APPROVAL FOR BIO-BEHAVIOURAL RESEARCH PROPOSAL INVOLVING HUMAN SUBJECTS AT AIISH

AIISH Institute Review Board (IRB)

Ref: SH/IRB/M.5/SLP.09/2025-26 **Date: 11.03.2026**

Title of the Dissertation	: Development and Validation of Hindi Auditory Naming Test-(HANT)
Name of the Student	: Ms. Chhavi Tyagi
Guide	: Dr. Ashishek B P
Co-Guide (if any)	: NIL
Proposed Duration of the Project	: 06 months
Source of funding	: Non-funded Master's dissertation
Estimated budget	: Nil
Reference Number of the Project	: Nil
Date on which the IRB meeting was held	: 05.03.2026
A clear statement of the decision reached IRB meeting (In the event of the proposal is not approved, a statement of reasons for the same must be indicated)	: **APPROVED**
Advice & suggestions (if any)	: NIL

Signature & Name of the Member Secretary-IRB
Dr. Animesh Barman
Professor of Audiology
All India Institute of Speech and Hearing, Mysore-570006

Member Secretary
Institutional Review Board
अखिल भारतीय वाक् श्रवण संस्थान
All India Institute of Speech and Hearing
मानसगंगोत्री / Manasagangothri
मैसूरु / Mysuru-570006

Phone : 0821-2502000 / 2502100 Toll Free No. 18004255218, Fax : 0821-2510515
e-mail : director@aiishmysore.in / @aiishmysuruofficial, @aiishmysuru
website : www.aiishmysore.in

APPENDIX II

Hindi Auditory Naming Test-(HANT)

S.no.	Stimulus	Expected Response
1.	कौन-सा जानवर रेगिस्तान की रेत में आसानी से चलता है? /kɔːn saː d͡ʒaːnʊər reːgɪsʈaːn kiː reːʈ̪meː aːsaːniː seː t͡ʃəlʈ̪aː hɛː/ Which animal walks easily on the desert sand?	ऊँट / Camel /uːʈ̃/
2.	कौन-सा फल बाहर से काँटेदार और अन्दर से बड़े बीजों वाला होता है? /kɔːn saː pʰəl baːhər seː kãːʈeːɖ̪aːr ɔːr ənɖ̪ər seː bəɽeː biːd͡ʒõː ʊaːlaː hoːʈ̪aː hɛː/ Which fruit is thorny outside and has big seeds inside?	कटहल/Jackfruit /kəʈʰəl/
3.	कौन-सी चीज़ चाबी से खुलती है? /kɔːn siː t͡ʃiːz t͡ʃaːbiː seː kʰulʈ̪iː hɛː/ Which thing opens with a key?	ताला / Lock /ʈ̪aːlaː/
4.	कौन-सा पक्षी लम्बे और सुन्दर पंखों वाला होता है? /kɔːn saː pəkʂiː ləmbeː ɔːr sunɖ̪ər pəŋkʰõː ʊaːlaː hoːʈ̪aː hɛː/ Which bird has long and beautiful feathers?	मोर / Peacock /moːr/
5.	घर के किस कमरे में खाना पकाया जाता है? /gʱər keː kɪs kəmreː meː̃ kʰaːnaː pəkaːjaː d͡ʒaːʈ̪aː hɛː/ In which room of the house is food cooked?	रसोई / Kitchen /rəsoːiː/
6.	सब्ज़ी काटने के लिए क्या इस्तेमाल करते हैं? /səbziː kaːʈneː keː lɪe kjaː ɪsʈ̪eːmaːl kərʈ̪eː hɛ̃ː/ What is used to cut vegetables?	चाकू / Knife /t͡ʃaːkuː/

7.	मरीज़ों का इलाज कौन करता है? /məriːzo͂ː kaː ɪlaːd͡ʒ kɔːn kərʈaː hɛː/ Who treats patients?	डॉक्टर / Doctor /ɖɔːkʈər/
8.	कौन-सा कीड़ा शहद बनाता है? /kɔːn saː kiːɽaː ʃəhəd bənaːʈaː hɛː/ Which insect makes honey?	मधुमक्खी / Honeybee /məɖʱuməkkʰiː/
9.	कौन-सी सब्ज़ी कड़वी और बाहर से खुरदरी होती है?	करेला / Bitter gourd
	/kɔːn siː səbziː kəɽʋiː ɔːr baːhər seː kʰurɖəriː hoːʈiː hɛː/ Which vegetable is bitter and rough on the outside?	/kəreːlaː/
10	कौन-सी सब्ज़ी लम्बी और सफ़ेद होती है?	मूली / Radish
	/kɔːn siː səbziː ləmbiː ɔːr səfeːɖ hoːʈiː hɛː/ Which vegetable is long, white, and a little bitter?	/muːliː/
11.	समय देखने के लिए दीवार पर जिसे टांगते हैं, उसे क्या कहते हैं?	घड़ी / Clock
	/səməj deːkʰneː keː lɪe ɖiːʋaːr pər d͡ʒɪseː ʈaːŋgʈeː hɛ̃ː, ʊseː kjaː kəhʈeː hɛ̃ː/ What is hung on the wall to see the time?	/gʱəɽiː/
12.	जो नदी के ऊपर लोगों के आने-जाने के लिए बनाया जाता है, उसे क्या कहते हैं?	पुल / Bridge
	/d͡ʒoː nəɖiː keː uːpər loːgõː keː aːneː d͡ʒaːneː keː lɪe bənaːjaː d͡ʒaːʈaː hɛː, ʊseː kjaː kəhʈeː hɛ̃ː/ What is built over a river for people to cross?	/pul/
13.	शरीर के किस अंग में अंगूठी पहनते हैं?	उँगली / Finger
	/ʃəriːr keː kɪs əŋg meː̃ əŋguːʈʰiː pəhənʈeː hɛ̃ː/ On which body part do we wear a ring?	/ũŋgliː/
14	पहाड़ से तेज़ी से गिरते हुए पानी को क्या कहते हैं?	झरना / Waterfall
	/pəɦaːɽ seː ʈeːziː seː gɪrʈeː ɦue paːniː koː kjaː kəhʈeː hɛ̃ː/ What is water falling quickly from a mountain called?	/d͡ʒʱərnaː/

15.	सुबह के समय घास और पत्तों पर जो छोटी-छोटी पानी की बूँदें दिखती हैं, उन्हें क्या कहते हैं? /subəɦ keː səməj gʱaːs ɔːr pət̪t̪õː pər d͡ʒoː t͡ʃʰoːʈiː t͡ʃʰoːʈiː paːniː kiː buːnd̪ẽː d̪ɪkʰt̪iː hẽː, ʊnʰẽː kjaː kəht̪eː hẽː/ What are the small drops of water seen on grass and leaves in the morning called?	ओस / Dew /oːs/
16.	पेंसिल से लिखा मिटाने के लिए क्या इस्तेमाल करते हैं? /peːnsɪl seː lɪkʰaː mɪʈaːneː keː lɪe kjaː ɪst̪eːmaːl kərt̪eː hẽː/ What is used to erase pencil writing?	रबर / Eraser /rəbər/
17.	बुखार होने पर डॉक्टर क्या देता है? /bukʰaːr hoːneː pər ɖɔkʈər kjaː d̪eːt̪aː hɛː/ What does a doctor give when a person has fever?	दवा /Medicine /d̪əʊaː/
18.	जो देश या राज्य को चलाती है उसे क्या कहते हैं? /d͡ʒoː d̪eːʃ jaː raːd͡ʒjə koː t͡ʃəlaːt̪iː hɛː ʊseː kjaː kəht̪eː hẽː/ What is the one who governs a country or a state called?	सरकार / Government /sərkɑːr/
19.	जब कोई अच्छी बात होती है तो हमें क्या महसूस होता है? /d͡ʒəb koːiː ət͡ʃʰiː baːt̪ hoːt̪iː hɛː t̪oː həmẽː kjaː məɦsuːs hoːt̪aː hɛː/ What do we feel when something good happens?	खुशी / Happiness /kʰuʃiː/
20.	बच्चे स्कूल में क्या करने जाते हैं जिससे उन्हें ज्ञान मिलता है? /bət͡ʃt͡ʃeː skuːl meⁿ kjaː kərneː d͡ʒaːt̪eː hẽː d͡ʒɪsseː ʊnʰẽː gjaːn mɪlt̪aː hɛː/ What do children go to school for, from which they gain knowledge?	पढ़ाई / Study /pəɽʱaːiː/
21.	जब अदालत अपराध साबित न होने पर व्यक्ति को सही ठहराती है, तो उसे क्या प्राप्त होता है? /d͡ʒəb əd̪aːlət̪ əpraːd̪ʱ saːbɪt̪ nə hoːneː pər ʊjəkt̪i koː səɦiː ʈʰəɦraːt̪iː hɛː, t̪oː ʊseː kjaː praːpt̪ hoːt̪aː hɛː/	न्याय / Justice /njaːj/

	When the court finds a person not guilty, what does he receive?	
22.	सोते समय जो दिखता है, उसे क्या कहते हैं? /soːt̪eː səməj d͡ʒoː d̪ɪkʰt̪aː hɛː, ʊseː kjaː kəht̪eː hɛ̃ː/ What is seen while sleeping called?	सपना / Dream /səpnaː/
23.	उम्र में छोटे लोगों को अपने से बड़ों को क्या देनी चाहिए? /umr meː̃ tʃʰoːʈeː loːgõː koː əpneː seː bəɽõː koː kjaː d̪eːniː tʃaːhɪeː/ What should younger people give elders?	इज़्ज़त / Respect /ɪzzət̪/
24.	घड़ी में क्या देखते हैं? /gʱəɽiː meː̃ kjaː d̪eːkʰt̪eː hɛ̃ː/ What do we see in a clock?	समय / Time /səməj/
25.	मेहनत करने से क्या मिलती है? /meːɦnət̪ kərneː seː kjaː mɪlt̪iː hɛː/ What do we get from hard work?	सफलता / Success /səpʰəlt̪aː/
26.	जब कोई आपको बार-बार परेशान करे तो आपको क्या आता है? /d͡ʒəb koːiː aːpkoː baːr baːr pəreːʃaːn kəreː t̪oː aːpkoː kjaː aːt̪aː hɛː/ What do you feel when someone troubles you again and again?	ग़ुस्सा / Anger /ɣʊssaː/
27.	जब कोई बच्चा किसी चीज़ को पाने के लिए बार-बार रोता, तो उसे क्या कहते हैं? /d͡ʒəb koːiː bət͡ʃt͡ʃaː kɪsiː t͡ʃiːz koː paːneː keː lɪe baːr baːr roːt̪aː, t̪oː ʊseː kjaː kəht̪eː hɛ̃ː/ What is it called when a child cries again and again to get something and insists on it?	जिद / Tantrum /zɪd̪/

28.	सच बोलने और सही काम करने की आदत को क्या कहते हैं? /sət͡ʃ boːlneː ɔːr səɦiː kaːm kərneː kiː aːɖ̪ət̪ koː kjaː kəhț̪eː hɛ̃ː/ What is the habit of speaking truth and doing the right thing called?	ईमानदारी / Honesty /iːmaːnɖ̪aːriː/
29.	रात के अँधेरे में अकेले होने पर हमें क्या लगता है? /raːt̪ keː ə̃ɖ̪ʱeːreː meː̃ əkeːleː hoːneː pər həmẽː kjaː ləgt̪aː hɛː/ What do we feel when we are alone in the darkness of night?	डर / Fear /ɖ̤ər/
30.	एक माँ अपने बच्चे से क्या करती है? /eːk maː̃ əpneː bət͡ʃt͡ʃeː seː kjaː kərt̪iː hɛː/ What does a mother do with her child?	प्यार / Love /pjaːr/

APPENDIX III

Response Form

S. No.	Expected response	Generated response	Score	Remark
1.				
2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10				
11.				
12.				
13.				
14				
15.				
16.				
17.				
18.				
19.				
20.				
21.				
22.				
23.				

24.				
25.				
26.				
27.				
28.				
29.				
30.				

Remarks: If correct/synonym, If partially correct (semantically related) , If incorrect (no response/incorrect response)

APPENDIX IV

Content Validation Form

To the evaluator,

I, Chhavi Tyagi, am pursuing my dissertation under Dr. Abhishek. B. P., Assistant Professor at the Centre of Speech-Language Sciences from All India Institute of Speech & Hearing, Mysuru. We are developing a clinician-administered assessment tool, namely, the Hindi Auditory Naming test (HANT). The present study aims to develop and validate an Auditory Naming Test in Hindi to assess lexical retrieval in the neurotypical population across age groups using auditory modality. Given your expertise in speech and language disorders, I would greatly appreciate your input in validating this tool. This validation ensures that the tool accurately captures the intended information and effectively serves its purpose. Your insights will be crucial in refining the tool for its eventual use. I have attached the test material below for your review. If you could kindly review the tool, I would be genuinely grateful. Feel free to share your feedback. Your expertise is highly respected, and I am eager to hear your thoughts. Thank you very much for considering my requests. Your contribution will have a significant impact on the quality of this tool.

This study involves neurotypical individuals aged 18–65 years to validate auditory stimuli and examine natural lexical access. The stimuli include high- and low-frequency nouns, both concrete and abstract, framed in simple to moderately complex sentences to avoid ceiling effects while allowing flexible and spontaneous responses.

Level 1 Objective:
To identify stimulus items with greater clarity and minimal ambiguity. Your task is to shortlist the stimuli based on these characteristics.

Level 2 Objective:
The test is designed to include items of both high and low frequency, as well as abstract and concrete items, in appropriate proportion. We request you to validate the content of the stimuli on these parameters.

Instructions: Please rate any number between 1-3 from the drop-down menu to validate each parameter.

- **Clarity of the stimulus items:** The item is written in a straightforward manner (1-Unclear;2-Somewhat clear;3-Very clear)

- **Ambiguity of the stimulus items:** The item is unclear and can be interpreted in more than one way.
- **Relevance of the stimulus items:** The item is culturally and contextually relevant (1-Completely relevant; 2-Somewhat relevant;3-Irrelevant)
- **Familiarity of the response items:** The response item is familiar (1- Not familiar; 2- Somewhat familiar; 3- very familiar)
- **Appropriateness of the response items:** The response accurately matches what the test intends to measure.

	Stimulus	***Expected Response***	***Clarity***	***Ambiguity***	***Relevance***	***Familiarity (of Response item)***	***Appropriateness (of Response item)***	***Suggestions***
1.	कौन सा जानवर म्याऊँ करता है? /ɦəm əpne: kɑ:no: se: kjɑ: kərt̪e: hɛ:/	बिल्ली /bɪl:i:/	***3-Very clear***	***very ambiguous***	***1 Completely relevant***	***3 very familiar***	Highly appropriate	Clear & Simple
2.	कौन सा फल लंबा और पीला होता है जिसे छीलकर खाते हैं? /ko:n sa: pʰəl ləmbɑ: ɔ:r pi:lɑ: ho:t̪a: hɛ: d͡ʒɪse: t͡ʃʰi:lkər kʰɑ:t̪e: hɛ:/	केला /ke:la:/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***3 very familiar***	***Highly appropriate***	Very Clear
3.	कौन सी चीज़ चाबी से खुलती है?	ताला /ta:la:/	***3 Very clear***		***1 Completely relevant***	***3 very familiar***		No changes

	/koːn siː t͡ʃiːz t͡ʃɑːbiː seː kʰʊl̪t̪iː hɛː/			***very ambiguous***			Highly appropriate	
4.	शरीर का तापमान किससे मापा जाता है? /ʃəriːr kɑː t̪ɑːpmaːn kɪsseː mɑːpɑː d͡ʒɑːt̪aː hɛː/	थर्मामीटर /tʰərmaːmiːʈər/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***2 somewhat familiar***	Highly appropriate	Some rural participants may find “थर्मामीटर” less familiar; still valid
5.	घर के किस कमरे में खाना पकाया जाता है? /gʰər keː kɪs kəmreː meːn kʰɑːnɑː pəkɑːjɑː d͡ʒɑːt̪aː hɛː/	रसोई /rəsɔːiː/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***3 very familiar***	Highly appropriate	Valid
6.	सब्ज़ी काटने के लिए क्या इस्तेमाल करते हैं? /səbziː kɑːʈneː keː lɪjeː kjɑː ɪst̪eːmɑːl kərt̪eː hɛː/	चाकू /tʃaːkuː/	***3 Very clear***	***very ambiguous***	***1 Completely relevant***	***3 very familiar***	Highly appropriate	Simple
7.	मरीज़ों का इलाज कौन करता है? /məriːzoːn kɑː ɪlɑːd͡ʒ koːn kərt̪ɑː hɛː/	डॉक्टर /ɖɔːkʈər/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	3 very familiar	Highly appropriate	Strong item

8.	कौन सा कीड़ा शहद बनाता है? /koːn saː kiːɽɑː ʃəhəd bənɑːʈaː hɛː/	मधुमक्खी /mədʱʊməkːʰiː/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***2 somewhat familiar***	Highly appropriate	Slightly less familiar word, but correct
9.	फ़ोन को किससे चार्ज करते हैं? /foːn koː kɪsseː t͡ʃɑːrd͡ʒ kərʈeː hɛː/	चार्जर /tʃaːrdʒər/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***3 very familiar***	Highly appropriate	
10.	बुखार होने पर डॉक्टर क्या देता है? /bʊkʰɑːr hoːneː pər ɖɔːkʈər kjɑː ɖeːʈaː hɛː/	दवा /dʊvə/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***3 very familiar***	Highly appropriate	
11.	समय देखने के लिए दीवार पर जिसे टांगते हैं, उसे क्या कहते हैं? / səmɛj ɖeːkʰne keː liːje ɖiːʋaːr pər d͡ʒɪseː ʈaːŋgʈe hɛ̃, ʊseː kjaː kəhʈe hɛ̃ /	घड़ी /gʱəɽiː/	***3 Very clear***	***Not ambiguous***	***3 Very clear***	***3 very familiar***	Highly appropriate	
12.	जो नदी के ऊपर लोगों के आने-जाने के लिए बनाया जाता है, उसे क्या कहते हैं? / d͡ʒoː nədiː keː ʊpər loːgoːn keː aːne-d͡ʒaːneː keː liːjeː bənaːjaː d͡ʒaːʈaː hɛ, ʊseː kjaː kəhʈeː hɛ̃ /	पुल /pʊl/	***3 Very clear***	***Not ambiguous***	***3 Very clear***	***3 very familiar***	Highly appropriate	

13.	जिसमें किताबें रखकर बच्चे स्कूल ले जाते हैं, उसे क्या कहते हैं? / ʤɪsme: kɪt̪a:be: rəkʰkər bətʃ:e sku:l le: ʤa:t̪e: hɛ̃, ʊse: kja: kəht̪e: hɛ̃ /	बस्ता /bʌsta:/	*3 Very clear*	*Not ambiguous*	*3 Very clear*	*3 very familiar*	Highly appropriate	
14.	एक ऐसा जानवर जो पानी और जमीन दोनों पर रहता है और जिसकी पीठ पर कठोर खोल होता है, उसे क्या कहते हैं? / e:k ɛ:sa: ʤa:nʊər ʤo: pa:ni: ɔ:r ʤəmi:n do:no: pər rɛhta: hɛ ɔ:r ʤɪski: pi:ʈʰ pər kəʈʰo:r kʰo:l ho:ʈa: hɛ, ʊse: kja: kəht̪e hɛ̃ /	कछुआ /kəʧʰʊa:/	*3 Very clear*	*Not ambiguous*	*3 Very clear*	*3 very familiar*	Highly appropriate	
15.	पहाड़ से तेज़ी से गिरते हुए पानी को क्या कहते हैं? / pəɦa:ɽ se: te:zi: se: gɪrt̪e: ʊe: pa:ni: ko: kja: kəht̪e: hɛ̃ /	झरना /ʤʱərna:/	*3 Very clear*	*Not ambiguous*	*3 Very clear*	*3 very familiar*	Highly appropriate	
16.	सुबह के समय घास और पत्तों पर जो छोटे-छोटे पानी की बूँदें दिखती हैं, उन्हें क्या कहते हैं? / sʊbəh ke: səməj gʱa:s ɔ:r pət̪t̪o: pər ʤo: ʧʰo:ʈe: ʧʰo:ʈe: pa:ni: ki:	ओस /o:s/	*3 Very clear*	*Not ambiguous*	*3 Very clear*	*2 somewhat familiar*	Highly appropriate	Some may find less familiar

	bu:nde: dɪ̱kʰt̪i̱: hɛ̃, ʊnhe: kja: kəht̪e: hɛ̃ /							
17.	पेंसिल से लिखा मिटाने के लिए क्या इस्तेमाल करते हैं? /pɛ:nsɪl se: lɪkʰa: mɪt̪ɑ:ne: ke: lɪje: kjɑ: ɪst̪e:mɑ:l kərt̪e: hɛ:/	रबड़ /rəbəɽ/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***3 very familiar***	Highly appropriate	
18.	जो देश या राज्य को चलाती है उसे क्या कहते हैं? /d͡ʒo: d̪e:ʃ ja: ra:d͡ʒjɐ ko: t͡ʃəla:t̪i̱: hɛ ʊsɛ kja: kəht̪e: hɛ̃/	सरकार /sərkɑ:r/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***3 very familiar***	Highly appropriate	Abstract, but good
19.	जब कोई अच्छी बात होती है तो हमें क्या महसूस होती है? :/d͡ʒəb ko:i ətt͡ʃʰi: ba:t̪ ho:t̪i̱: hɛ t̪o: həmɛ̃ kja: məɦəsu:s ho:t̪i̱: hɛ/	ख़ुशी /kʰʊʃi:/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***3 very familiar***	Highly appropriate	
20.	बच्चे स्कूल में क्या करने जाते हैं जिससे उन्हें ज्ञान मिलता है /bət͡ʃ:e sku:l me: kya: kərne d͡ʒa:te: hɛ̃ d͡ʒɪsse: ʊnhe: gja:n mɪlta: hɛ?/	पढ़ाई /pəɽʱa:i:/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***3 very familiar***	Highly appropriate	
21.	जब अदालत अपराध साबित न होने पर व्यक्ति को सही ठहराती है, तो उसे क्या प्राप्त होता है?	न्याय /nʲɑ:j/	***3 Very clear***		***1 Completely relevant***	***3 very familiar***		Higher-order concept, but important

	/dʒəb ədɑ:lət əprɑ:dʱ sɑ:bɪt nʌ ho:ne pər vəktɪ ko səɦi: ʈʰəhra:ti: hɛ, to: u:se /nʲɑ:j/ prɑ:pt ho:tɑ: hɛ?/			***Somewhat ambiguous***			Highly appropriate	
22.	सोते समय जो दिखता है, उसे क्या कहते हैं? /so:ʈe: səmaj d͡ʒo: dɪkʰʈa: hɛ ʊse: kja: kəhʈe: hɛ̃/	सपना /səpna:/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***3 very familiar***	Highly appropriate	
23.	उम्र में छोटे लोगों को बड़ों की क्या करनी चाहिए? /ʊmr me:n ʧʰo:ʈe: lo:go:n ko: bəɽo:n ki: kjɑ: kərni: ʧɑ:hi:je:/	इज़्ज़त /ɪz:ət/	***3 Very clear***	***Somewhat ambiguous***	***1 Completely relevant***	***2 somewhat familiar***	Highly appropriate	Abstract but they might say “Sewa” so bit ambigous
24.	घड़ी में क्या देखते हैं? /gʰəɽi: me: kja: ɖe:kʰʈe: hɛ̃/	समय /səməj/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***3 very familiar***	Highly appropriate	
25.	मेहनत करने से क्या मिलती है? /me:ɦənət̪ kərne: se: kja: mɪlʈi: hɛ/	सफलता /səfəlta:/	***3 Very clear***	***very ambiguous***	***1 Completely relevant***	***3 very familiar***	Highly appropriate	Higher Cognition
26.	चोट लगने पर जो तकलीफ होती है, उसे क्या कहते हैं?	दर्द /dərd/	***3 Very clear***		***1 Completely relevant***	***3 very familiar***		Clear

	/ʧoːʈ̪ ləgneː pəɾ d͡ʒoː ʈ̪əkliːf hoːʈ̪iː hɛ ʊseː kjaː kəhʈ̪eː hɛ̃/			***Not ambiguous***			Highly appropriate	
27.	मुसीबत में किसी की सहायता करने को क्या कहते हैं? /musiːbət meː kisiː kiː səhaːjtaː kərneː koː kjaː kehəteː hɛ/	मदद /mədəd/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***3 very familiar***	Highly appropriate	Strong
28.	सच बोलने और सही काम करने की आदत को क्या कहते हैं? /səʧ boːlneː ɔːr səɦiː kaːm kərneː kiː aːd̪əʈ̪ koː kjaː kehəteː hɛ/	ईमानदारी /iːmaːndaːriː/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***2 somewhat familiar***	Highly appropriate	Its Quite abstract but acceptable
29.	जब कोई व्यक्ति बहुत नाराज़ होता है, उसे क्या आता है?" /d͡ʒəb koiː ʊyakti bəhʊt naːraːz hɔːtaː hɛ, ʊseː kjaː aːtaː hɛ/	गुस्सा /gʊsːaː/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***3 very familiar***	Highly appropriate	Clear
30.	थक जाने के बाद इंसान घर जाकर क्या करता है? /ʈ̪ʰək d͡ʒaːne keː baːd ɪnsaːn kjaː kərtaː hɛ/	आराम /aːraːm/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***3 very familiar***	Highly appropriate	Valid

31.	कौन सा जानवर चपटा और लंबी पूँछ वाला है, उसकी त्वचा मोटी और खुरदरी होती है, और यह पानी और जमीन दोनों पर रहता है? /kaːʊn saː d͡ʒaːnʊər t͡ʃəpʈaː ɔːr ləmbiː pũːt͡ʃ ʊaːlaː hɛ, ʊskiː ʈʊət͡ʃaː moːʈiː ɔːr kʰʊrɖriː hoːʈiː hɛ, ɔːr jəh paːniː ɔːr d͡ʒəmɦiːn doːnoː pər rəhʈaː hɛ?/	मगरमच्छ /mʌgərməʧʰ/	***3 Very clear***	***Not ambiguous***	***1 Completely relevant***	***2 Useful but not essential***	Highly appropriate	Question is Slightly long, but descriptive

Overall rating of the Test items

If you have any other comments/ suggestions, Please mention them below.

- Suggestions for rephrasing or revising the question to improve clarity

- What do I feel is there must be a balance between Abstract and concrete items chosen as target word
- Culturally the stimuli are very strong but simple to moderate complex sentences will be more suitable. keep sentence length moderate as item no 14 and 31 are very long, Need to be Rephrase.
- Few target words chosen, there might be a regional variability from urban and rural areas. Please check for it.

- Comment on the overall structure and organization of the questionnaire

Excellent sequencing of questions from Basic to functional and abstract concepts. I feel it is well structured for both linguistic and cognitive load. All the best for this awesome work. It is need of this hour.

APPENDIX V

5.a. Informed consent form

I have been informed about the procedures of the study. The possible risks and benefits of my participation as a human subject in the study are clearly understood by me. I understand that I have a right to refuse participation as a subject or withdraw my consent at any time. I am also aware that by subjecting this investigation, I must give more time for assessments by the investigating team and that these assessments may not benefit me. I have the freedom to write to Chairman AEC, in case of any violation of these provisions without the danger of me being denied any rights to secure the clinical services at this institute.

I, _____ the undersigned, give my consent to be a

participant in this investigation/study/program.

Name and signature of the Participant/Caretaker:
Date:

Name and signature of Investigator (Student):
Date:

5.b. सहमति पत्र

मुझे इस अध्ययन/शोध की पूरी प्रक्रिया के बारे में जानकारी दी गई है। मुझे यह भी समझाया गया है कि इसमें भाग लेने से मुझे क्या लाभ और क्या जोखिम हो सकते हैं। मैं समझता/समझती हूँ कि मुझे इस अध्ययन में शामिल न होने या बीच में ही अपनी सहमति वापस लेने का पूरा अधिकार है।

मुझे यह भी बताया गया है कि इस अध्ययन में भाग लेने पर मुझे जाँच करने वाली टीम को थोड़ा ज़्यादा समय देना होगा और यह जाँचें मेरे लिए सीधे तौर पर फायदेमंद नहीं भी हो सकती हैं।

अगर मुझे लगे कि इन नियमों का पालन नहीं किया जा रहा है, तो मुझे ए.ई.सी के चेयरमैन को लिखने की पूरी स्वतंत्रता है, और ऐसा करने से मुझे इस संस्थान में मिलने वाली क्लिनिकल सेवाओं से वंचित नहीं किया जाएगा।

मैं, _____, नीचे हस्ताक्षर करने वाला/वाली, अपनी इच्छा से इस अध्ययन/कार्यक्रम में भाग लेने की सहमति देता/देती हूँ।

प्रतिभागी/अभिभावक का नाम व हस्ताक्षर:

तारीख:

जांचकर्ता

(छात्र) के हस्ताक्षर: