Unveiling the Diversity: A Systematic Review of the Famous Faces Tests

Demi van den Ende, 7271689

MSc Neuroscience and Cognition, UU

June 14, 2023

Supervisors:

Dr. R. Mark, Assistant professor[1]

E. van den Elzen, PhD-candidate[1]

Examinor:

Prof. Dr. M.J.E. van Zandvoort[2]

[1] *Department of Cognitive Neuropsychology, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, the Netherlands*

[2] *Utrecht University, Utrecht, the Netherlands*

Abstract

Background: The lack of a standardized protocol for the Famous Face(s) Test has led to a wide range of variations across studies. These include differences in the stimuli (e.g. type, number), tasks used to assess performance (e.g. recall, recognition), and many more factors, all, making it difficult to compare results. Establishing a standardized protocol for the Famous Faces Test would be helpful not least because it has been found (in four previous studies) to be sensitive to detect preclinical Alzheimer's disease. The current systematic review explored differences and similarities between the Famous Faces Tests used in studies conducted over the past five years in an attempt to make recommendations for standardizing the protocol for this test.

Methods: This systematic review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) guidelines. In April 2023, four databases (PubMed, PsychINFO, Web of Science and Embase) were systematically searched using specific keywords. Studies were selected based on predetermined in- and exclusion criteria that would include studies with performance based measures on the Famous Faces Test in individuals with cognitively normal aging and those along the continuum of Alzheimer's disease. A quality assessment was created to explore and asses the Famous Faces Tests used in the included studies across multiple domains, including stimuli characteristics, stimulus selection, and measurements.

Results: In total, 11 studies were included. Variations between FFTs used in these studies on all three domains were apparent. Main differences in stimuli characteristics were the number of stimuli, stimulus duration and control measure,

whereas main differences for stimulus selection were the details of reporting, and finally, within measurements, main differences existed in type of tasks and scoring method.

Discussion: This review revealed the many differences in the FFTs used across studies underscoring the necessity of a standardized protocol/FFT. This could then be used in the clinic and perhaps compared to what is currently used in diagnostic practice to detect (pre)clinical dementia in the earliest stages.

Keywords: cognitive assessment, dementia, Famous Faces Test, memory, systematic review

Layman summary

Achtergrond: De Famous Faces Test is een test die het geheugen meet voor bekende gezichten. Er is momenteel geen gestandaardiseerd protocol (vast format) voor de Famous Faces Test, wat ertoe heeft geleid dat er veel verschillende variaties zijn in stimuli (zoals de soort, het aantal), taken, structuur, uitvoering en beoordeling. Het is hierdoor moeilijk om resultaten tussen studies te vergelijken. Vier onderzoeken hebben tot nu toe laten zien dat de Famous Faces Test mogelijk kan bijdragen bij de diagnose van Alzheimer in een vroeg stadium. Het is daarom cruciaal om een gestandaardiseerd protocol op te stellen. Deze review heeft geprobeerd de eerste stap te zetten om dit doel te bereiken door de bestaande variaties in recent gebruikte Famous Faces Tests te onthullen.

Methode: Deze systematische review werd uitgevoerd aan de hand van de PRISMA-richtlijnen, wat kan worden gezien als een checklist. In april 2023 zijn vier databases (PubMed, PsychINFO, Web of Science en Embase) doorzocht met

behulp van specifieke zoekwoorden. Studies zijn geselecteerd op basis van vooraf bepaalde eisen, waarin werd vastgesteld welke studies wel en niet zouden worden meegenomen. Studies werden meegenomen als ze uitkomsten gaven van het presteren op de test en moest ofwel gezonde ouderen onderzoeken, of mensen die zich in een van de stadia van de ziekte van Alzheimer bevinden. Er werd een kwaliteitsbeoordeling gemaakt om te kijken naar hoe goed elke studie de Famous Faces Test omschrijft, zodat deze in de toekomst door anderen op dezelfde manier kan worden uitgevoerd. De resultaten van deze review werden samengevat op basis van drie domeinen: kenmerken van stimuli, selectie van stimuli en metingen.

Resultaten: In totaal zijn 11 studies meegenomen in de review. In elk van de drie domeinen zijn verschillen tussen de testen gevonden. Wat betreft de kenmerken van de stimuli waren de belangrijkste verschillen het aantal stimuli, de duur van de stimuli en de controlemaatregel. Wat betreft de stimulusselectie waren de belangrijkste verschillen de details van het rapporteren. En tot slot bestaan er binnen de metingen belangrijke verschillen in het soort taken en de manier van scoren.

Discussie: Deze review onthulde de verschillen in studies met betrekking tot kenmerken van stimuli, selectie van stimuli en metingen, waaruit blijkt dat een protocol volgens een vast format noodzakelijk is. Door richtlijnen op te stellen voor de selectie van stimuli, controlemaatregelen en taakuitvoering kunnen de resultaten makkelijker worden vergeleken tussen studies, wat mogelijk kan helpen bij de diagnose van de ziekte van Alzheimer in de vroege fase.

Table 1. Definitions of keywords and concepts used throughout the review

Term	Description
Alzheimer's disease (AD) continuum	Describes the progression of AD from its asymptomatic phase to its most severe form of dementia. Along this continuum, individuals transition through different phases of AD, including preclinical AD, mild cognitive impairment (MCI) due to AD, and mild, moderate or severe dementia due to AD.[1,2]
Categories	Refers to different ways (including occupations) in which the famous personalities are known e.g., athletes, singers, TV personalities, politics.
Cued recall	A type of memory retrieval that involves using cues to access stored information. These cues are often related to the meaning of the information, such as its semantic category or associated words.[3]
Declarative memory	A type of long-term memory that involves conscious remembering of facts and events. It is also known as explicit memory, because it requires conscious effort to retrieve the information. Declarative memory can be divided into two subtypes: episodic memory and semantic memory.[4,5]
Episodic memory	Allows/Enables us to recall personal experiences and events that are tied to a specific time and place.[6] Complements semantic memory by adding a subjective and temporal dimension to our memories.[7]
Free recall	A type of memory retrieval where a person has to produce previously learned information without any cues or hints.[3] In recall, a person can only rely on recollection. For example, a person performing a recall task has to come up with the person's name, therefore they cannot rely on familiarity and must rely on recollection.
Identification	A type of memory retrieval which requires semantic knowledge about the person, such as their profession or other relevant information.[8]
Recognition	The ability to identify an item as previously experienced. It is believed to involve two processes: familiarity and recollection. Familiarity is when the subject sees a person's face and feels a vague sense of having met the person before. Recollection is when the subject sees a person's face and remembers specific details of the encounter.[9] In recognition, a person can rely on familiarity in addition to, or instead of, recollection. E.g., if a person performing a recognition task already knows the name before looking at the options, they rely on recollection. But if they can also rely on familiarity by looking at the options.

Remote memory	The ability to recall events or information from the distant past, such as years or decades ago.[10,11]
Semantic memory	Enables us to recall general knowledge such as facts, concepts, and ideas, that is not tied to a specific time or place.[6]
Temporal gradient	Referred to as a temporal gradient for memory loss, that is thought to be greater for recent events compared to remote events (Ribot's law).[10]
Time of prominence	Refers to the time at which the famous personality became or was famous e.g. certain decades.
Tip-of-the-tongue (TOT) phenomenon	The experience of having a strong feeling that information is stored in one's memory, but failing to retrieve.[12,13]

Introduction

The Famous Faces Test (FFT) is a neuropsychological assessment tool designed to evaluate the ability to recognize and name famous people from photographs, such as celebrities, politicians or athletes. During a FFT participants are shown photographs of famous people, typically from different decades, and are asked to name them or provide information about them. One of the FFTs' first applications was in 1971 by Sanders and Warrington,[10] who researched retrograde memory defects in five amnesia patients. The presentation of famous faces from different time periods (decades) made it possible to research the extend of remote memory impairment and to explore if a temporal gradient of memory loss exists, e.g., greater memory loss for recent events compared to remote events. Since then, the test has been widely used in clinical and research settings to evaluate memory impairments in a broad range of patients e.g., Korsakoff syndrome,[14] epilepsy,[15] and neurodegenerative disorders like Parkinson's disease[16] and Alzheimer's disease (AD).[17]

Recently, there has been an increasing focus on investigating the FFT's utility for the early diagnosis of AD.[18–20] For a clinical AD diagnosis, a significant cognitive decline in at least one domain (e.g., memory, language or executive function) that interferes with daily functioning and cannot be explained by other causes is required.[21] However, AD pathology begins to develop decades before the onset of clinical symptoms. The concept of the AD continuum describes the progression of AD from its asymptomatic phase to its most severe form of dementia,[1,2] along this continuum individuals transition through different phases of AD, including preclinical AD, mild cognitive impairment (MCI) due to AD, and mild, moderate or severe dementia due to AD. The preclinical phase of AD poses diagnostic challenges since

individuals may exhibit AD-related brain pathology without noticeable cognitive impairment at a behavioural level. Currently, preclinical AD can only be identified via AD biomarkers, including amyloid plaques and neurofibrillary tangles, for example, which often require neuroimaging or cerebrospinal fluid analysis.[22] However, these methods are not easily accessible and require costly and invasive procedures. Moreover, the presence of biomarkers does not guarantee that a person will develop AD, as half of the people with positive biomarkers will not show clinical symptoms of AD in their lifetimes.[23] Given that preclinical AD presents a critical window for intervention and prevention, as current treatments can only slow down the symptom progression and not reverse it,[24] it is essential to explore alternative methods to detect AD in its preclinical phase.

Estévez-González et al[19] conducted a study to investigate the ability to detect memory impairment for famous faces, which is a common symptom of dementia due to AD, earlier along the AD-continuum. The results showed that two years prior to clinical AD diagnosis, patients performed more poorly on the FFT compared to patients with MCI and cognitively normal controls. Another large study[12] showed the FFTs' potential (a temporal gradient for memory loss (i.e. greater loss for recent memories compared to remote memories) even earlier on the continuum, in patients with preclinical AD and amyloid-pathology. These findings demonstrated the FFTs' ability to aid in distinguishing between patients who would later further progress to AD, even at a time no behavioural changes are evident, suggesting its potential contribution to preclinical AD diagnosis.

The FFT involves presenting participants with photographs of famous personalities from different categories (e.g., celebrities, politicians or athletes) and different decades (e.g. 1970s, 2000s), asking them to provide the name belonging to

that face and/or to provide other information about the personality. Performance on the FFT relies on prior knowledge that does not require encoding or retrieval strategies.[25] It employs both semantic and episodic memory, components of our declarative long-term memory involving conscious remembering of facts and events.[4] Semantic memory enables us to recall general knowledge such as facts, concepts, and ideas, without being tied to a specific time or place. On the other hand, episodic memory allows us to recall personal experiences and events that are associated with a specific time and place.[6] A debate has emerged regarding whether the FFT primarily taps into semantic or episodic memory. While the majority of researchers believe that the FFT can be used to assess semantic memory, it should be noted that no test is entirely 'pure' in terms of exclusively targeting a single cognitive function.[26]

Numerous discrepancies exist among studies concerning FFT components, primarily due to the lack of a standardized protocol. Variations can be observed not only in the stimuli used but also in how they are presented and how performance is assessed across studies. For instance, some studies incorporate a greater number of famous face stimuli compared to others, and while some employ black-and-white photographs, others utilize coloured photographs. Furthermore, different tasks are employed. In a recall task, for example, participants must name the famous personality, while in a recognition task they typically need to select which of the photographs represents the famous personality. Variations are evident in the number of tasks, their order of presentation, and the types of tasks included (e.g., recall, recognition). Occasionally, studies vary in the names the assign to their tasks. Consequently, it remains unclear whether a recognition score, for instance, assigned in one study, can be directly compared to a recognition score from another study.

Moreover, the scoring methods employed can also differ among studies. Some studies calculate total accurate scores, while others consider errors and attempt to elucidate their nature, such as investigating whether they stem from a tip-of-the-tongue (TOT) experience or from a genuine lack of knowledge.[10,12,15,17,25]

Objective: The absence of a standardized protocol in the FFTs has led to a wide range of variations in stimuli, tasks, structure, administration, and scoring method, making it difficult to compare results across studies. Considering the potential use of FFT in preclinical AD diagnosis, it's crucial to establish a 'gold standard'. An initial step towards achieving this goal involves unveiling the existing variations and key differences between FFTs. This review aimed to address this issue by examining recent studies and providing insights into the main discrepancies and similarities observed in FFTs.

Methods

This systematic review adhered to the PRISMA Statement (Preferred Reporting Items for Systematic Review and Meta-Analysis),[27] which is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses. The PRISMA 2020 Statement consists of a checklist of 27 items and a three-phase flow diagram that guide the reporting of the review process and the results. The checklist covers essential elements for transparent and complete reporting of a systematic review.

Search strategy

In April 2023, a systematic literature search was performed using four databases (PubMed, PsychINFO, Web of Science and Embase), using a combination of keywords and MeSH terms related to famous faces, Alzheimer's

disease and older adults. The search strategy, outlined in Table 1 of the Appendix, was agreed upon by R. Mark (R.M.), an assistant-professor and E. van den Elzen (E.E.), a PhD-candidate from Tilburg University who are both experts in this research domain. The author screened the titles and abstracts of all retrieved articles and excluded those that did not meet the inclusion criteria. The remaining articles were sought for availability and eligibility, and were included in the review if they met all of the inclusion criteria and none of the exclusion criteria. Three domains were checked for all studies, including stimuli characteristics, stimulus selection, and measurement.

Study selection

For this systematic review, the inclusion criteria were as follows: 1) Population: patients on the AD continuum (i.e., preclinical AD - MCI - Alzheimer's Dementia) or older adults; 2) Journal articles, review articles, dissertations; 3) Cognitive tests that assess famous faces; 4) Performance based measures.

Studies were excluded for further analysis if they contained: 1) Other populations (e.g., Parkinson's disease, Huntington Disease, Epilepsy or different age groups) or other types of Dementia (e.g., Frontotemporal Dementia and Semantic Dementia); 2) Case studies, books, citations, conference proceedings, talks and posters; 3) Cognitive test that assess famous names, voices, places and/or events 4) Neuroimaging results only.

Data

Information regarding the FFT that included stimuli characteristics, stimulus selection, and measurement, was collected to provide an extensive overview of the FFTs used in studies over the last five years. Additionally, information about the purpose of the included studies, samples of interest and relevant findings regarding

the FFT were collected. In instances where information was missing, these gaps were duly noted as ‘NA’ (not available), as this proved valuable for the overall assessment of the FFT tests included. The abovementioned information was systematically organized and presented in Tables 2 and 3.

Task interpretation

Given the absence of a golden standard for the FFT, studies may differ in their naming and interpretation of a task. To ensure clarity within the scope of this review and to facilitate accurate interpretation of findings across studies, the following (task) terms/names/types were used: 1) **Recall**, this task involves presenting participants with famous face stimuli and subsequently asking them to name the face in the photograph, one at a time. Recall can be either 'free recall,' where participants must attempt to remember the famous person's name without any cues, or 'cued recall,' where participants are provided with a cue (e.g., the sound of the first two letters of the name) to assist in recalling the name. 2) **Identification**, this task requires participants to demonstrate semantic knowledge about the person, such as their profession or other relevant information. 3) **Recognition**, typically a forced-choice format, in which participants are presented with a set of stimuli and must indicate whether they know the person, or a multiple-choice format to select the famous face amongst a set of distractor faces. If studies deviated from the tasks outlined above or introduced tasks that could not be attributed to any of these task categories, an explanation was provided in Table 3.

Quality assessment

To assess the quality of the included studies’ FFT a quality assessment tool was created. Other existing quality assessment tools were considered first, however, a decision was made to create our own because no previous quality assessment tool

would fit the data of this review well enough (e.g., tools from the Joanna Briggs Institute[28]). For this review, the importance lied in exploring the similarities and differences of the studies' FFTs, in order to later on work towards a gold standard of the FFT. Therefore, this quality assessment compared stimulus and task-characteristics of the FFT used in the 11 studies included here.

The quality assessment used a cumulative point system, whereby each reported item received 1 point and each non-reported item received 0 points. Within thirteen items, the following three domains were assessed: Stimuli characteristics, Stimulus selection and Measurement. A maximum of eight points was possible for Stimuli characteristics, one point for Stimulus selection, and four points for Tasks and measurement. Subsequently, descriptive analyses were conducted on the collected data, and the total scores were transformed into z-scores using Microsoft Excel. Based on the deviation from the mean, the z-scores were then converted into a quality rating. Z-scores greater than 1 SD above the mean were classified as Good quality, while z-scores falling within the range of -1 to 1 SD were classified as Fair quality, and z-scores less than 1 SD below the mean were classified as Poor quality studies. This quality assessment process was reviewed and approved by the same scientists who validated the search strategy (R.M. and E.E.). The results of the quality assessment are represented in Table 4.

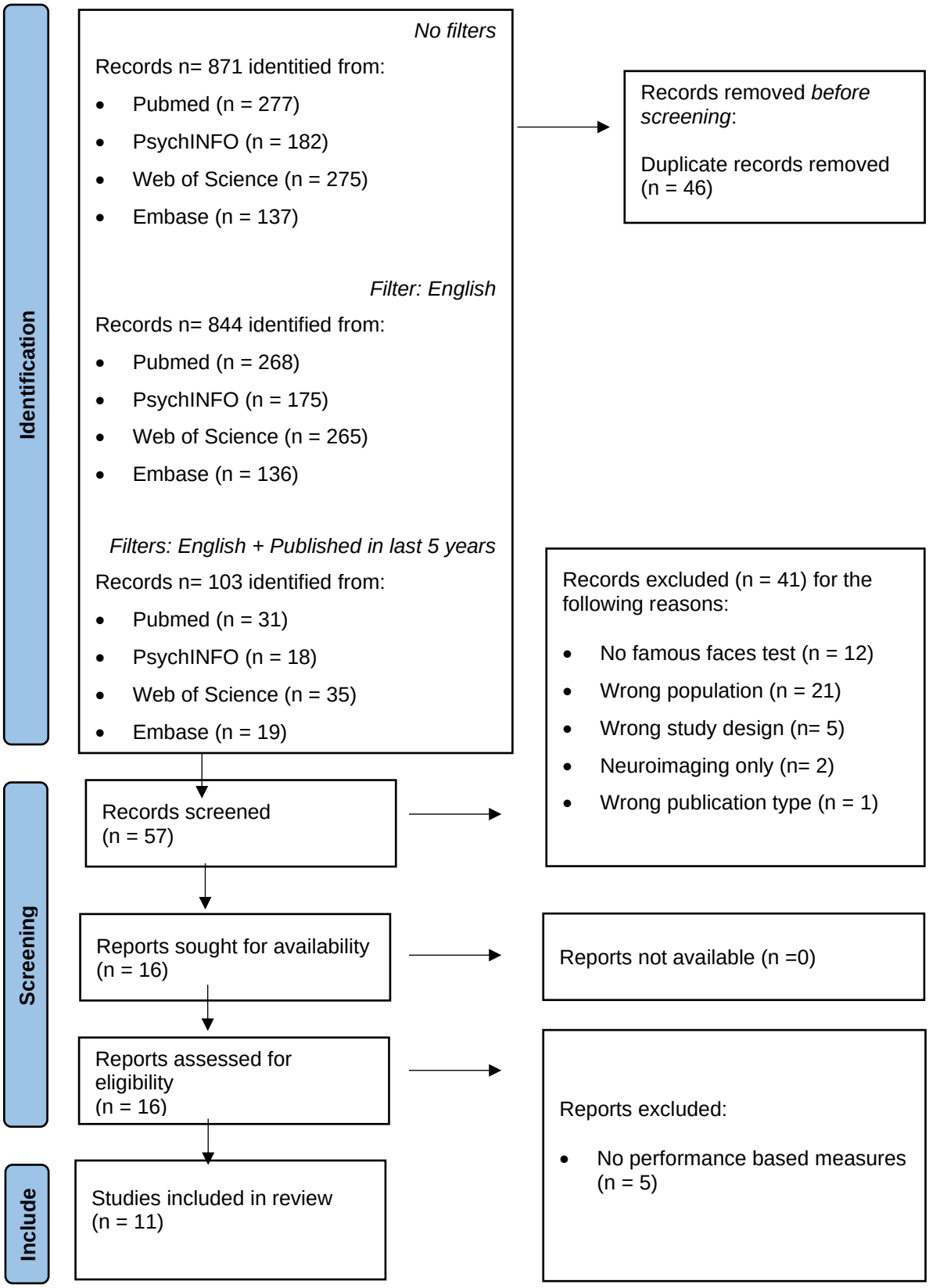
Results

Search results and study selection

The search strategy was part of a larger review study, so it was initially performed without any filters. This yielded 871 relevant studies for further screening, of which 844 were in English. Due to the guidelines established by Utrecht University

for this assignment, the scope of this review was limited to studies conducted in the past five years. This reduced the number of relevant studies to 103. After eliminating 46 duplicates, 57 unique studies were left for further screening. Based on their title and abstract, 41 studies did not meet the predetermined inclusion criteria and were excluded. The remaining 16 studies were available for retrieval. Of the retrieved reports, 5 did not meet inclusion criteria,[29–33] as they did not provide performance based measures. In the end, 11 studies remained and were included in this systematic review (Figure 2). Results were synthesized based on three domains: Stimuli characteristics, Stimulus selection and Measurements. The collected information was organized in Table 2, which presents a summary of the included studies, and Table 3 which presents an extensive overview of the FFT characteristics.

Figure 2. Flow diagram of search and study selection



[a] As the search strategy was part of a larger review study, a complete overview of the search strategy is presented here. However, for the current review the scope was limited to studies from the last five years only. Figure was adapted from [27]

Table 2. Summary of studies included in the review

Year	Author	Purpose	Sample of interest	Relevant findings
2018	Orlovsky et al[25]	to better understand whether famous face naming exhibited a relationship with PET amyloid outcomes similar to a traditional measure of semantic memory (i.e. object naming).	125 clinically normal older adults (90 Aβ - and 35 Aβ +)	◦ Generally better recall for more remote famous faces compared to recent famous faces, regardless of the cued recall condition. ◦ Consistent with previous research suggesting a decline in semantic memory along the AD trajectory during the prodromal and preclinical stages of the disease.
2018	Pineault et al[34]	to investigate the brain mechanisms associated with semantic performance in aMCI patients	14 aMCI and 14 healthy older adults	◦ Response times were longer and less accurate for aMCI patients than healthy controls. ◦ No correlation between the participants' performance on the FFT and their performance on any of the more classical neuropsychological tasks used to assess semantic memory.
2019	Borghesani et al[35]	to evaluate the contribution made by semantic task-related fMRI measures to predicting AD pathophysiology or dementia.	20 AD-spectrum; 18 NCs	◦ AD patients' overall performance was lowest for the recall task, compared to recognition tasks, but no significant results were found for AD patients.
2019	Nanda et al[8]	to assess visual associative learning and famous face recognition ability among subjects with stable amnestic MCI relative to early stage dementia due to AD and cognitively normal healthy controls (NC) and to correlate these differences with volumetric changes on MRI	22 MCI; 19 AD; 20 NC	◦ Scores differed significantly between MCI and controls, and the AD and MCI cohort, with AD scoring lower than MCI and NCs, and MCIs scoring lower than NCs.
2019	Rahmani et al[36]	to study and compare naming or recognizing famous faces with the recognition of newly learned faces among patients with Amnesia Mild Cognitive Impairment and Alzheimer's Disease.	63 healthy older adults; 60 aMCI; 62 AD	◦ Patients with AMCI and Alzheimer's disease obtained significantly lower scores than those in the control group on both recall and recognition tasks. ◦ The updated FFT is highly sensitive to early cognitive changes relevant to facial recall and recognition.

2020	Campos-Magdaleno et al[37]	to identify longitudinal patterns of TOTs in individuals with subjective complaints and with MCI regarding progress of their cognitive status.	193 SCC; 33 MCI stable; 23 MCI worsened	◦ Phonological access (TOTs) differentiated MCI patients, stable and worsened, from adults with SCCs at all evaluation times. ◦ phonological access declined over time in the three groups, without significant interactions between groups and time.
2020	Kelly & McDonald[38]	to develop a brief, comprehensive screen for social difficulties suitable for older adults, that can be administered alongside other cognitive measures. The initial step in this development was to examine the construct validity of the new instrument.	28 healthy adults and 26 dementia patients	◦ The dementia group performed significantly worse on the Face Identification task (assessing free recall and recognition) compared to the healthy adults.
2021	García et al[39]	to investigate if the Famous Faces Naming Test can help to identify which patients will progress to Alzheimer's disease and who will not.	8 MCI-AD; 17 MCI-non AD	◦ MCI patients who converted to AD 2 years later performed significantly worse on Famous Faces Naming Test (they named less and described more items) compared to MCI patients who did not convert over that time period.
2021	Tennant et al[40]	to investigate the ability of healthy older adults to retrieve the names of famous faces in relation to positron emission tomography measurements of amyloid-β plaques and tau neurofibrillary tangles.	85 healthy older adults	◦ Proper name retrieval was correlated with the episodic memory composite ($r = 0.30$, $p = 0.002$) and logical memory ($r = 0.32$, $p = 0.003$) ◦ No sex differences in proper name retrieval or episodic memory.
2022	Souza et al[41]	to establish culturally adapted norms for proper names' pictures from an adult sample, in psycholinguistic measures (naming and categorization scores) and evaluative dimensions (fame, familiarity, distinctiveness, arousal, and representational quality).	56 younger adults; 51 older adults	◦ Age effects were observed across all variables, except familiarity. Familiarity was relatively immune to aging. Older participants rated the pictures as more arousing, famous, distinctive, and with higher representational quality than younger ones.
2023	Van den Elzen et al[42]	to explore the possibility of developing a Famous Faces Test specifically for older adults living in the Netherlands to distinguish preclinical AD from normal aging.	188 older adults (more participant were involved in the	◦ The systematic collection and selection of famous faces through five steps led to nine short D-FFT versions of equivalent mean difficulty. These nine D-FFT versions allow for the comparison of cognitively

	to explore the possibility of a test which could distinguish preclinical AD from normal aging.	stimulus selection process)	normal older adults and clinical populations by means of a short, enjoyable task that can be performed at a time and location convenient for the participant.

Abbreviations: AD, Alzheimer's Disease; aMCI, amnestic Mild Cognitive Impairment; BASS-D, Brief Assessment of Social Skills – Dementia; D-FFT, Dutch-Famous Faces Test; DK, ''don't know'' response; fMRI, functional magnetic resonance imaging; MCI-AD, patients with MCI which converted to AD 2 years later; MCI-non-AD, patients with MCI which did not convert to AD 2 years later; MRI, magnetic resonance imaging; NC, (cognitively) normal controls; PET, positron emission tomography; PN, proper name; SCC, subjective cognitive complaints; SM, semantic memory; TOTs, tip-of-the-tongue experiences.

Table 3. Overview of famous face stimuli, selection method and measurements

Reference	Number	Presentation	Categories	Prominence	Stimulus selection	Tasks	Scoring
Orlovsky et al[25]	24	◦ Computer or tablet ◦ Coloured photographs ◦ White background ◦ Max 20 sec	TV, sports, radio, and cinema	1960-1980 and 1990-present	Piloted on healthy and cognitively normal older adults, not enrolled in the present study, to ensure that chosen faces were familiar and well known.	1. Free recall: provide a name and occupation (identification) 2. Cued recall: phonemic cues were provided by giving the sound of the first two letters of the famous figure's first name.	◦ Overall total of correctly recalled remote and recent famous faces without and with a phonemic cue. ◦ A response was quantified as correct if the participant recalled the first or last name of the famous face.
Pineault et al[34]	96	◦ Computer screen ◦ Coloured photographs ◦ Black background ◦ Equivalent in size	Athletes, singers, actors or political figures	1930 - present	NA	1. Identification: a semantic category (e.g., 'actor') was presented and participant had to decide whether each subsequent stimuli belonged to that category. This included a matching a nonmatching condition.	◦ Reaction times and number of correct responses were averaged for each participant. ◦ Items left unanswered were recorded as incorrect responses.
Borghesani et al[35]	60	◦ Computer screen ◦ Black and white photographs	Different professional categories	NA	A free recall task on 20 normal male subjects (age range 18–33 years). From 200 stimuli, only those faces that were named within the 5 s by at	1. Free recall: naming 2. Identification: match two famous faces on profession (3 MC)	◦ Successful naming ◦ Successful matching ◦ Successful selection

		◦ No background cues ◦ Matched for mean luminance			least 19 subjects were included.	3. Recognition: select the famous face amongst distractor faces (4 MC)	
Nanda et al[8]	15	◦ Computer screen ◦ 15 sec	National politicians, sportspersons, musicians, film actors and social activists.	Range of generations, mostly before 2000	Selected from the Internet.	○ Recognition: select the famous face amongst distractor faces (3 MC) ○ Free recall: name	Total score: ◦ 1 point if participants correctly recognized the famous face but were unable to name them, and 2 points if they were able to name them also. ◦ 0 points if they were unable to recognize the famous face or in case no answer was proved by the allotted time.
Rahmani et al[36]	25	NA	NA	NA	Tested on 40 cognitively healthy people (mean age: 41) to ascertain appropriate difficulty. Accuracy scores in recognizing ranged from 85% to 100% (M = 97.4; S.D. = 2.84).	1. Free recall: provide a name and description (identification)	Average total scores and errors scores: ◦ 2 points for correctly recalling the name (either full name or only the first or second name). ◦ 2 points when at least 2 pieces of information were provided, but only 1 point if 1 detail was given. This included only details

								which could not be interpreted from the picture itself, such as sex.
Campos-Magdaleno et al[37]	50	◦ Computer screen ◦ Coloured photographs	National and international actors, singers, politicians, sportsmen, arts personalities, etc.	Last 50 years	Performance measures of 20 cognitively unimpaired participants (age range: 55-80) were used to select stimuli from a pool of 70. The final 50 stimuli had the highest punctuation in familiarity and semantic information (age, residence, marital status etc.).	1. <u>Free recall</u>: indicate name familiarity by speaking the name, or choosing "I don't know," or “I can’t recall the name at the moment” (TOT). The photographs that produced TOT responses, were presented again in attempt to name them. 2. <u>Identification</u>: if the TOT persisted or an incorrect name was given, participants answered questions about the person and their name, such as their profession. 3. <u>Recognition</u>: select the correct name from two non-target names (3 MC), and indicate if it matched the attempted recall Additionally participant rated the familiarity of 50 stimuli on a 1-5 scale (1 = unfamiliar, 5 = highly familiar).	Four sub-scores: A. Subjective familiarity: calculated by summing the familiarity responses for all 50 photographs. B. Feeling of Knowing: the number of times that participants indicated they knew the name. C. Semantic access: successful access to the semantic representations of the names [(CORs + pTOTs)/N]. D. Phonological access: the proportion of both successful semantic and phonological retrievals [CORs/(CORs + pTOTs)]

Kelly & McDonald[38]	16	NA	NA	1930s through 1970s	Stimuli were piloted and the 16 most readily recognized were selected.	Three questions as follows: 1. Recognition: do you recognize this person (yes/no) 2. Free recall: what is their name? 3. Identification: what are they known for?	◦ Each questions was scored 1 (correct) or 0, yielding a score of 3 per item, with a maximum score of 48.
García et al[39]	22	◦ A4-sized folder ◦ Coloured photographs ◦ Randomized order	Singers, politicians, kings, actors, TV personalities, celebrities of gossip shows and sportsmen	1950s and recent	NA	1. Free recall: naming 2. Identification: in case of incorrect naming or omissions, participants were instructed to provide a detailed description of the famous person.	1. Naming score: the number correctly recalled names (last name or full name) without cues. 2. Identification score: the number of items producing specific identifying information, usually after an omission 3. TOT score: calculation NA
Tennant et al[40]	20	◦ Printed ◦ Black and white photographs	Cultural icons as entertainers, politicians, or international acknowledged leaders	75% from 1970 or later, with peak fame from 1920 to current	Selected on: 1) popularity and celebrity status in visual media and press, 2) race and sex, and 3) time epoch in which the person was famous. Tested on: 30 cognitively healthy individuals (mean age: 45) with >90%	1. Free recall: naming 2. Identification: if the participant failed to name the person, they were asked to cite semantic facts about the person.	Accuracy of both naming and recognition: ◦ if the first and last name could be produced within 5 sec, 2 points were given for naming and recognition ◦ if only the first name was produced, 1 point for naming

			(65% male).		accuracy on naming and recognition			and 2 points for recognition were given ◦ Maximum 2 points per famous face for free recall or if failed, maximum 2 points for recognition
Souza et al[41]	40	◦ Mix of grey-scale and coloured photographs ◦ Blank background ◦ Randomized order ◦ Resized at 500X500 pixels ◦ Controlled for luminosity	Culture, entertainment and TV, sports, leaders (50% male).	1940s and current	Three native Portuguese selected 120 stimuli, of which two judges provided the correct target and category names and evaluated the appropriateness (IRA of 86.67%). Disagreements were discussed with a third judge and excluded if no consensus was reached or it was considered of lower cultural relevance.	1. Rating: familiarity, fame, distinctiveness, and arousal 2. Free recall: naming, or choose "do not know" (DK) or "know the item but are momentarily unable to name it" (TOT). Including identification, in case of TOT responses, where participants could provide semantically related information. 3. Identification: choose the best category (4 MC with an "I don't know" option presented in the end) 4. Representational quality rating: assessing the potential of each picture in representing the concept/name.	◦ Name accuracy (%), categorization accuracy (%), modal name (the most referred valid name) and name agreement (% of agreement regarding the modal name). ◦ The respective value of name variability was estimated to capture the conceptual variability in correctly naming the item. ◦ Errors were incorrect names, and incorrect response included Errors (%) together with DK (%) and TOT (%).

						◦ The 40 famous face stimuli were combined with 40 well-known places stimuli.		
Van den Elzen et al[42]	20	◦ PC, tablet, or smartphone ◦ Black and white photographs ◦ White background ◦ 15 sec ◦ Resized to 250 by 333 pixels	Film & theatre; music; politics; sports	1960s; 1970s; 1980s; 1990s; 2000s; 2010s	Famous peoples' names from 1940 to 2020 were selected and tested for recognition by Dutch older adults. If ≥ 30% recognized the names, photographs where provided and tested using free recall. A total of 220 open-source photographs was further divided into 10 different FFT versions with equivalent mean difficulty.	1. <u>Free recall :</u> participants were asked if they knew the person's name and could choose from the responses: "Yes", "Yes, but it is on the tip of my tongue", "No, but I recognize the face," or "No, and I do not recognize the face." If they answered "Yes," they were asked to enter the name of the person. 2. <u>Recognition:</u> choose the correct name that fits the photograph (MC).	◦ The cumulative number of famous faces per proportion of correct naming responses to the 20 famous faces in each FFT version. ◦ (Recognition scores were not further explained, as they were not included in the selection process this paper was about).

Abbreviations: CORs, correct responses; IRA, interrater agreement; NA, not available; pTOT, positive TOT (i.e. participants can describe the famous person, although they cannot recall the name); RMA; repeated measures analysis; TOT, tip-of-the-tongue.

Quality appraisal of the data

Based on the quality assessment presented in Table 4, three studies[34,41,42] can be regarded as exemplifying good quality in terms of their reporting of the FFT. These studies showed a commendable level of detail and transparency regarding the characteristics of the FFT, the method used for stimulus selection, and the measurements. Conversely, six studies were classified as fair quality, denoting that they provided a satisfactory amount of information regarding their FFT. However, when compared to the good-quality studies, these fair-quality studies exhibited some gaps or omissions in their reporting.[8,25,35,37,39,40] Lastly, two studies were categorized as poor quality due to their inadequate documentation of essential information concerning the FFT, which distinguished them from the other included studies.[36,38]

Table 4. Quality assessment of the studies' FFTs

Studies		25	34	35	8	36	37	38	39	40	41	42
Stimuli characteristics	Number	1	1	1	1	1	1	1	1	1	1	1
	Colour	1	1	1	0	0	1	0	1	1	1	1
	Background	1	1	1	0	0	0	0	0	0	1	1
	Size	0	1	0	0	0	0	0	0	0	1	1
	Luminance	0	0	1	0	0	0	0	0	0	1	0
	Stimulus duration	1	0	0	1	0	0	0	0	0	0	1
	Categories	1	1	1	1	0	1	0	1	1	1	1
	Time of prominence	1	1	0	1	0	1	1	1	1	1	1
Stimulus selection	General description	1	0	1	0	1	1	1	0	1	1	1
Measurement	Tasks	1	1	1	1	1	1	1	1	1	1	1
	Practice trial	0	1	0	0	0	0	0	0	0	1	1
	Feedback	0	1	0	0	0	0	0	1	0	0	0
	Total	8	9	6	5	3	6	4	6	6	10	10
	Quality	Fair	Good	Fair	Fair	Poor	Fair	Poor	Fair	Fair	Good	Good

[a] Using the newly created quality assessment, each study was evaluated on Stimuli characteristics, Stimulus selection and Measurement, where 1 point could be received for each item.

[b] The cumulative scoring system was as follows: for each item, if information was provided within the study itself, or if a correct reference containing this information was provided, one point was received. If no information was provided, the item received 0 points. Items included: Stimuli characteristics: a) Number: the number of stimuli; b) Coloured: coloured or black-and-white photographs; c) Background: information about the background and/or background cues; d) Size: size of the photographs; e) Luminance: luminance of the photographs; f) Stimulus duration: the

duration the stimulus was presented; g) Categories: categories of the famous persons were indicated e.g., media or politics; h) Time of prominence: the time at which the famous persons became or was famous e.g., certain decades. Stimulus selection: a) General description: information about how the stimuli were selected, e.g., based on performance of a control group. Measurement: a) Tasks: information about all the tasks that were included in the study's version of the FFT; b) Practice trail: information about whether or not a practice trial was included; c) Feedback: information about whether or not information about the performance was given.

[c] ∗To convert the total scores into quality ratings, raw scores were first converted into z-scores. Z-scores > 1 were considered of **Good quality** (raw scores ≥ 9), while z-scores from -1 to 1 were considered **Fair quality** (raw scores from 5-8), and z-scores < 1 were considered **Poor quality** studies (raw scores ≤ 4).

Stimuli characteristics

The total number of stimuli in each FFT differed between studies (M = 25.27, SD = 24.89), ranging from 15 to 96 ([8],[34], respectively). Three studies used grey-scale photographs,[35,40,42] while four studies used coloured photographs,[25,34,37,39] and others did not provide information about this. Out of the 4 studies that did not indicate whether they used grey or coloured photographs,[8,36,38,41] one study publicly shared their set of stimuli which contained a mix of grey-scale and coloured photographs.[41] In two studies, the photographs were controlled for luminance,[35,41] while three studies matched the stimuli on size,[34,41,42] and four corrected for background cues by adding a blank background colour to remove any background cues.[25,34,35,42] Figure 1 present an example of the stimuli, corrected for background cues.

Figure 1. Example of Famous Face stimuli

Adapted from [27]

Except for two studies,[36,38] all studies indicated what the famous personalities were famous for (referred to as 'categories' in Table 3). Some studies chose to use famous personalities from different occupational[34,35] or media categories,[25] while six studies included personalities from multiple fields,[8,37,39–42] e.g. singers, politicians, kings, actors, TV personalities, and sportsmen. One study[42] specifically described how these categories were divided across FFTs. The majority of the studies did not mention if national and/or internal personalities were used as stimuli. However, [8] mentioned to only have included national figures, while two studies used both national and international figures.[37,41] Furthermore, 82% of the studies specified the time at which the famous personalities were or became famous.[8,25,35,37–43] Of these studies, three further divided time of prominence into two time-categories,[25,39,41] one subdivided it into three time-categories,[43] and another used six decades as different time-categories.[42] Finally, two studies did not provide information about time of

prominence.[35,36] Apart from three studies,[8,34,39] all studies described their method of selecting the famous face stimuli.

In more than half of the studies, the stimuli were presented on a computer screen or laptop,[8,34,35,37,42–44] while 18% of the studies printed the stimuli on paper.[39,40] The remaining three studies did not indicate how the stimuli were presented.[36,38,41] Only five studies indicated how long each stimulus was presented, which ranged from 500 milliseconds to 20 seconds (M = 11.10, SD = 8.05).[8,25,34,42,43] Furthermore, five studies indicated presenting the stimuli in a randomized order,[34,39,41,43,44] of which three studies counterbalanced blocks, categories, or conditions.[34,41,43]

Stimulus selection

The method of stimuli selection was described in 8 studies (73%),[25,35–38,40–42] providing information about, for example, their control sample, the task that was used to determine which stimulus had to be selected, and performance measures, all increasing reproducibility. There were two studies[34,39] that did not describe stimulus select at all, and one that only mentioned selecting stimuli from the internet.[8]

Measurement

The FFTs were all differently constructed and varied in the type of task they used to assess performance. Based on the task distinction previously described, 90% of the studies included a free recall task where participants were presented the famous face stimuli and were asked to provide a name.[8,25,35–42] From these studies, three also included options that would indicate the participant does not know the name or TOT experience.[37,41,42] Besides the free recall task, one study[25] additionally added a cued recall task, proving a phonemic cue. An identification task was

included in 82% of the studies.[25,34–41] This was implemented in various ways, such as a multiple-choice format,[41] a matching format,[34,35] or a describing or answering questions (about) the personalities.[25,36–38,40] Recognition tasks were included in 45% of the studies, which could either be a forced-choice (yes or no),[38,42] or multiple-choice format.[8,35,37,42] In three studies, a practice trial proceeded the experimental trails.[34,41–44] Finally, one study added a quality rating task, to assess the representational quality of the stimuli.[41] No study mentioned providing the participants with feedback on their performance. Two studies specifically indicated not providing feedback on the performance,[34,39] although in Pineault et al[34] participants received feedback about whether each answer was recorded.[8]

Finally, the scoring method differed between studies. The most common scoring method, was to calculate a cumulative response of total correct responses.[25,35,37,39,42] However, two studies used the average of correct responses,[36,45] while three others[8,39,40] used a predetermined point system in which they specified the points that could be received, e.g., if only the first name was produced, this resulted in 1 point for recall and 2 points for recognition. Pineault et al[34] also measured reaction times, and six studies further examined TOT experiences, in which an option such as ‘the name is on the tip of my tongue’ was provided.[37,37,39–42] Of all measurement methods, Souza et al[41] used the most extensive method of scoring, calculating percentages for total correct responses, errors and incorrect responses.

Discussion

This review explored studies published in the last five years that used an FFT. The aim was to compare and contrast the FFTs as an initial step in working toward a standardized protocol. This study exposed differences and similarities in the FFTs

across studies on all three domains. The diversity observed in the FFTs across studies provides valuable insights into the variations in stimulus characteristics, selection methods, and measurements. These diversities raise important considerations for future research and the development of a standardized protocol.

Stimuli characteristics

One aspect of diversity in the FFT is the variation in the number of stimuli. Studies utilized a wide range of stimuli, indicating a lack of consensus on the optimal number. This variability may impact the reliability and generalizability of the results. The use of different photograph types, such as grey-scale and coloured, further contributes to the diversity. This variation may introduce bias and influence participants' responses, as colour could enhance memory performance, as shown by Spence et al.[46] In their study a recognition task was performed on neutral scenes, in either colours or grey-scale. They found that recognition was better in the coloured condition than in the grey-scale condition. While several studies acknowledged the importance of standardizing stimuli, including, for example, luminance, size, and background cues, there was a lack of consistency in their application across the literature. Notably, even the pioneering study by Sanders and Warrington[10] failed to account for these crucial components. However, in the present review of more recent studies, it is encouraging to note that researchers have begun to try to standardize the stimuli they use for these factors, recognizing their potential influence on performance. These advancements signify a positive trend towards improved experimental precision and provide a stronger basis for understanding the impact of stimuli characteristics on cognitive tasks.

The famous personalities in the FFT varied in terms of the categories represented. Some studies focused on specific occupational or media categories, while others included personalities from multiple fields. The use of national and/or international figures as stimuli was not consistently specified, introducing variability in familiarity and recognition across participant populations. Furthermore, the time of prominence, indicating when famous personalities achieved fame, varied across studies. This temporal dimension is particularly relevant in this field of research as it enables the investigation of memory loss extent reflecting a temporal gradient. Notably, studies exploring memory impairments, such as those associated with AD, have demonstrated that recent memories are often more severely affected while remote memories remain relatively intact. This temporal gradient phenomenon was found in the first studies that included a FFT,[10,14] and was more recently confirmed by, for example, the Harvard Aging Study who researched healthy aging.[12] However, in the current literature, the approaches to categorizing time of prominence differ, limiting cross-study comparisons.

The mode of stimulus presentation varied, with most studies utilizing computer screens or laptops. Some studies printed stimuli on paper, while a few did not specify the presentation method. Computerized assessment holds significant advantages for cognitive testing, as it allows for precise measurements of response times and eliminates potential human errors or subjective interpretations. Additionally, the scoring process becomes more streamlined and efficient as computerized systems can automatically score responses based on predefined criteria. Another advantage is the standardized administration of tests, ensuring consistent presentation and instructions across participants, thereby minimizing variations that could impact test performance. These benefits collectively underscore

the value of computerized assessment in cognitive testing and highlight its potential to improve the reliability and objectivity of cognitive evaluations.[47] According to Fillit et al,[48] such tests are even accessible for people with cognitive impairment or of older ages. Furthermore, information on stimulus duration and order was limited, and randomized presentation order and counterbalancing techniques were used in some studies.

Stimulus selection and measurement

Studies varied widely in terms of details reported on their stimulus selection method. Whereas many studies provided information about their control group, the tasks used to select the stimuli, and results of performance difficulty or agreement measures, which enhanced reproducibility of their FFT, some studies did not describe their selection method at all.

In terms of task diversity, the majority of the tasks could be attributed to either recall, identification, or recognition as previously defined. These three types of tasks rely on distinct cognitive processes. For example, in a recognition task, a person can rely on familiarity in addition to, or instead of, recollection, while in a recall task, a person can only rely on recollection. These different processes may vary in difficulty, as was indicated by e.g.,[49], where recollection was impaired on the FFT, but familiarity was unimpaired. Three studies included a practice trial before the experimental trials, which allowed participants to familiarize themselves with the task and stimuli. The inclusion of a practice trial helps ensure that participants understand the task requirements. One study incorporated a quality rating task to assess the representational quality of the stimuli. This task provided an additional dimension of evaluation, allowing researchers to determine the perceived quality and

representativeness of the selected stimuli. However, this process was typically already accounted for in the stimuli selection process. None of the studies mentioned providing feedback to the participants regarding their performance, of which two studies explicitly stating not providing feedback. One study only gave feedback regarding the recording of each answer.

The scoring methods employed in the FFTs varied across studies. Most of them calculated the cumulative response of the total correct responses, while some used the average or a point system based on the accuracy of recall or recognition. Assessing the speed of participants' responses was not very common, as only one study measured reaction times. More than half of the studies assess TOT experiences. While pioneering studies did not assess these TOT experiences,[10,17] the amount of details of measurement have shown to increase over the years, by searching for explanations as to why participants provide an incorrect answer. For example, in the study with the most extensive measurement method,[41] percentages for total correct responses, errors, and incorrect responses were calculated, providing a more comprehensive assessment of performance.

Limitations

In this review, we adhered to specific inclusion criteria to ensure a focused examination of recent studies. Consequently, the scope of our review was intentionally limited, as we exclusively considered studies published within the past five years. Although this approach allowed us to capture recent implementations, it is acknowledged that certain pivotal FFT studies were not encompassed in our review. As a consequence, the opportunity to investigate the evolutionary trajectory of the many FFTs over time was missed. Incorporating these pioneer studies would have provided valuable insights into the foundational aspects of these tests, shedding light

on their origin, development, and subsequent modifications. Recognizing this limitation, it is worth acknowledging that the exclusion of earlier studies does not dismiss their significance within the field. These pioneering works laid the groundwork for subsequent research and played a pivotal role in establishing the conceptual framework and empirical basis for the FFT. Future investigations may consider incorporating a historical perspective, encompassing these foundational studies, to provide a comprehensive overview of the FFT's evolution.

Another noteworthy limitation of this review is that it primarily focused on identifying differences amongst versions of the FFT rather than examining variations in performance outcomes across these tests. While the analysis shed light on the discrepancies in test characteristics and methodologies, it did not explore the potential disparities or possible similarities in performance among different patient groups. Of interest for future research would be to examine how performance on the FFT differs between patient groups in high-quality studies compared to fair- or poor quality studies. Assessing performance differences between these groups would provide valuable insights into the robustness and validity of the test.

Future research

In order to achieve a more comprehensive understanding of the FFT and its evolution over time, future research should aim to replicate the current study while expanding its scope to include a historical overview of the test, and highlighting the adaptations and modifications it has undergone. Building upon the findings of such a comprehensive overview, the next step would involve creation a standardized protocol for the FFT. This protocol would allow for a systematic approach that could be implemented across different countries, ensuring the selection of stimuli that best align with each cultural background. To ensure the quality and uniformity of the

stimuli used in the FFT, specific guidelines should be established. These guidelines should address various criteria, such as controlling for background cues, ensuring consistent luminance and size, and employing a fixed duration for presenting the stimuli. Additionally, decisions should be made regarding the use of grayscale or colored photographs, with clear criteria outlined for their selection. Furthermore, it is crucial for future iterations of the FFT to incorporate a consistent set of tasks in a standardized order. Based on the findings of this review, it is recommended to include a recall, identification, and recognition task, presented in a randomized order and counterbalancing between blocks of stimuli. This standardized task protocol would enable researchers to compare results more accurately across different studies, enhancing the validity and reliability of the test. By implementing these recommendations, future research on the FFT can contribute to the establishment of a robust and standardized framework. This framework would facilitate cross-study comparisons, promote cultural sensitivity in stimuli selection, and enhance the reliability and validity of the test.

In conclusion, this review revealed discrepancies in FFTs regarding stimuli characteristics, stimulus selection, and measurements, underscoring the necessity of a standardized protocol. By establishing guidelines for stimuli selection, control measures, administration, and scoring, the results can be effectively compared across studies, potentially aiding in preclinical Alzheimer's disease diagnosis.

References

1. Aisen PS, Cummings J, Jack CR, et al. On the path to 2025: understanding the Alzheimer's disease continuum. *Alz Res Therapy*. 2017;9(1):60. doi:10.1186/s13195-017-0283-5

2. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*. 2011;7(3):280-292. doi:10.1016/j.jalz.2011.03.003

3. Moult M. Cued Recall. In: Kreutzer JS, DeLuca J, Caplan B, eds. *Encyclopedia of Clinical Neuropsychology*. Springer; 2011:751-752. doi:10.1007/978-0-387-79948-3_1116

4. Riedel WJ, Blokland A. Declarative Memory. In: Kantak KM, Wettstein JG, eds. *Cognitive Enhancement*. Vol 228. Handbook of Experimental Pharmacology. Springer International Publishing; 2015:216. doi:10.1007/978-3-319-16522-6_7

5. Declarative Memory. Published November 3, 2022. Accessed June 4, 2023. https://www.simplypsychology.org/declarative-memory.html

6. Irish M, Piguet O. The Pivotal Role of Semantic Memory in Remembering the Past and Imagining the Future. *Frontiers in Behavioral Neuroscience*. 2013;7. https://doi.org/10.3389/fnbeh.2013.00027

7. APA Dictionary of Psychology. Accessed June 4, 2023. https://dictionary.apa.org/

8. Nanda S, Mohanan N, Kumari S, et al. Novel Face-Name Paired Associate Learning and Famous Face Recognition in Mild Cognitive Impairment: A Neuropsychological and Brain Volumetric Study. *Dement Geriatr Cogn Dis Extra*. 2019;9(1):114-128. https://doi.org/10.3758/s13428-022-01823-6

9. Merkow MB, Burke JF, Kahana MJ. The human hippocampus contributes to both the recollection and familiarity components of recognition memory. *Proc Natl Acad Sci U S A*. 2015;112(46):14378-14383. doi:10.1073/pnas.1513145112

10. Sanders HI, Warrington EK. Memory for remote events in amnesic patients. *Brain*. 1971;94(4):661-668.

11. Dorrego MF, Sabe L, García Cuerva A, et al. Remote Memory in Alzheimer's Disease. *JNP*. 1999;11(4):490-497. doi:10.1176/jnp.11.4.490

12. Huijbers W, Papp KV, LaPoint M, et al. Age-Related Increases in Tip-of-the-tongue are Distinct from Decreases in Remembering Names: A Functional MRI Study. *Cereb Cortex*. Published online August 30, 2016. doi:10.1093/cercor/bhw234

13. Brown AS. A review of the tip-of-the-tongue experience. *Psychological Bulletin*. 1991;109(2):204-223. doi:10.1037/0033-2909.109.2.204

14. Albert MS, Butters N, Levin J. Temporal Gradients in the Retrograde Amnesia of Patients With Alcoholic Korsakoff's Disease. *Archives of Neurology*. 1979;36(4):211-216. doi:10.1001/archneur.1979.00500400065010

15. Seidenberg M. Recognition and identification of famous faces in patients with unilateral temporal lobe epilepsy. *Neuropsychologia*. 2002;40(4):446-456. doi:10.1016/S0028-3932(01)00096-3

16. Leplow B, Dierks C, Herrmann P, Pieper N, Annecke R, Ulm G. Remote memory in Parkinson's disease and senile dementia. *Neuropsychologia*. 1997;35(4):547-557. doi:10.1016/S0028-3932(96)00116-9

17. Greene JDW, Hodges JR. Identification of famous faces and famous names in early Alzheimer's disease: Relationship to anterograde episodic and general semantic memory. *Brain*. 1996;119(1):111-128. doi:10.1093/brain/119.1.111

18. Werheid K, Clare L. Are Faces Special in Alzheimer's Disease? Cognitive Conceptualisation, Neural Correlates, and Diagnostic Relevance of Impaired Memory for Faces and Names. *Cortex*. 2007;43(7):898-906. doi:10.1016/S0010-9452(08)70689-0

19. Estévez-González A, García-Sánchez C, Boltes A, et al. Semantic Knowledge of Famous People in Mild Cognitive Impairment and Progression to Alzheimer's Disease. *Dement Geriatr Cogn Disord*. 2004;17(3):188-195. doi:10.1159/000076355

20. Arango-Lasprilla JC, Cuetos F, Valencia C, Uribe C, Lopera F. Cognitive changes in the preclinical phase of familial Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*. 2007;29(8):892-900. doi:10.1080/13803390601174151

21. American Psychiatric Association, ed. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5-TR*. Fifth edition, text revision. American Psychiatric Association Publishing; 2022.

22. Jack CR, Holtzman DM. Biomarker modeling of Alzheimer's disease. *Neuron*. 2013;80(6):1347-1358. http://dx.doi.org/10.1016/j.neuron.2013.12.003

23. Alzheimer's Association. 2023 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*. 2023;19(4):1598-1695. doi:10.1002/alz.13016

24. Dubois B, Hampel H, Feldman HH, et al. Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimers Dement*. 2016;12(3):292-323. doi:10.1016/j.jalz.2016.02.002

25. Orlovsky I, Huijbers W, Hanseeuw BJ, et al. The relationship between recall of recently versus remotely encoded famous faces and amyloidosis in clinically normal older adults. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*. 2018;10(1):121-129. doi:10.1016/j.dadm.2017.11.003

26. Salthouse TA. Selective review of cognitive aging. *J Int Neuropsychol Soc*. 2010;16(5):754-760. doi:10.1017/S1355617710000706

27. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews*. 2021;10(1):89. doi:10.1186/s13643-021-01626-4

28. JBI's Tools Assess Trust, Relevance & Results of Published Papers: Enhancing Evidence Synthesis | JBI. Accessed June 8, 2023. https://jbi.global/critical-appraisal-tools

29. Anderson AJ, Lin F. How pattern information analyses of semantic brain activity elicited in language comprehension could contribute to the early identification of Alzheimer's Disease. *NeuroImage Clin*. 2019;22:101788. doi:10.1016/j.nicl.2019.101788

30. Mazzi C, Massironi G, Sanchez-Lopez J, De Togni L, Savazzi S. Face Recognition Deficits in a Patient With Alzheimer's Disease: Amnesia or Agnosia? The Importance of Electrophysiological Markers for Differential Diagnosis. *Front Aging Neurosci*. 2020;12:580609. doi:10.3389/fnagi.2020.580609

31. Jonin PY, Duché Q, Bannier E, et al. Building memories on prior knowledge: behavioral and fMRI evidence of impairment in early Alzheimer's disease. *Neurobiol Aging*. 2022;110:1-12. doi:10.1016/j.neurobiolaging.2021.10.013

32. Tinard S, Guillaume F. Age-Related Differences in the Impact of Prior Knowledge on Recognition Performance: A Face Recognition Study. *Exp Aging Res*. 2019;45(2):154-166. doi:10.1080/0361073X.2019.1586108

33. Youm A, Moscovitch M. Aging, pattern separation, and categorical perception of faces. *Neuropsychologia*. 2021;161. https://doi.org/10.1016/j.neuropsychologia.2021.107999

34. Pineault J, Jolicoeur P, Grimault S, et al. Functional changes in the cortical semantic network in amnestic mild cognitive impairment. *Neuropsychology*. 2018;32(4):417-435. doi:10.1037/neu0000466

35. Borghesani V, Narvid J, Battistella G, et al. "Looks familiar, but I do not know who she is": The role of the anterior right temporal lobe in famous face recognition. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*. 2019;115:72-85. https://doi.org/10.1016/j.cortex.2019.01.006

36. Rahmani F, Fathi M, Kazemi M, Bahadori E. Recognition of famous and unfamiliar faces among patients suffering from amnesia mild cognitive impairment (AMCI) and Alzheimer's disease. *Iran J Psychiatry*. 2019;14(3):227-235.

37. Campos-Magdaleno M, Leiva D, Pereiro A, et al. Longitudinal Patterns of the Tip-of-the-Tongue Phenomenon in People With Subjective Cognitive Complaints and Mild Cognitive Impairment. *FRONTIERS IN PSYCHOLOGY*. 2020;11. doi:10.3389/fpsyg.2020.00425

38. Kelly M, McDonald S. Assessing social cognition in people with a diagnosis of dementia: Development of a novel screening test, the Brief Assessment of Social Skills (BASS-D). *JOURNAL OF CLINICAL AND EXPERIMENTAL*

NEUROPSYCHOLOGY. 2020;42(2):185-198. doi:10.1080/13803395.2019.1700925

39. García S, Cuetos F, Novelli A, Martínez C. Famous faces naming test predicts conversion from mild cognitive impairment to Alzheimer's disease. *Acta Neurol Belg*. 2021;121(6):1721-1727. doi:10.1007/s13760-020-01483-3

40. Tennant VR, Harrison TM, Adams JN, La Joie R, Winer JR, Jagust WJ. Fusiform gyrus phospho-tau is associated with failure of proper name retrieval in aging. *Annals of Neurology*. 2021;90(6):988-993. doi:10.1002/ana.26237

41. Souza C, Carmo JC, Garrido MV. Norms for pictures of proper names: contrasting famous people and well-known places in younger and older adults. *Behav Res*. Published online May 27, 2022. doi:10.3758/s13428-022-01823-6

42. van den Elzen EHT, Brehmer Y, Van Deun K, Mark RE. Stimulus material selection for the Dutch famous faces test for older adults. *Front Med (Lausanne)*. 2023;10:1124986. doi:10.3389/fmed.2023.1124986

43. Kalscheur EJ, Kandah CC, Woodard JL, Song W, Seidenberg M. Duration of Fame and Extent of Semantic Knowledge of Famous Names in Cognitively Intact Older Adults. *ARCHIVES OF CLINICAL NEUROPSYCHOLOGY*. 2019;34(8):1382-1391. doi:10.1093/arclin/acy109

44. Pistono A, Busigny T, Jucla M, et al. An Analysis of Famous Person Semantic Memory in Aging. *EXPERIMENTAL AGING RESEARCH*. 2019;45(1):74-93. doi:10.1080/0361073X.2018.1560118

45. Younes K, Borghesani V, Montembeault M, et al. Right temporal degeneration and socioemotional semantics: Semantic behavioural variant frontotemporal dementia. *Brain: A Journal of Neurology*. 2022;145(11):4080-4096. https://doi.org/10.1093/brain/awac217

46. Spence I, Wong P, Rusan M, Rastegar N. How color enhances visual memory for natural scenes. *Psychological Science*. 2006;17(1):1-6.

47. Gould CE, Ma F, Loup JR, Juang C, Sakai EY, Pepin R. Chapter 24 - Technology-based mental health assessment and intervention. In: Hantke N, Etkin A, O'Hara R, eds. *Handbook of Mental Health and Aging (Third Edition)*. Academic Press; 2020:401-415. doi:10.1016/B978-0-12-800136-3.00024-7

48. Fillit HM, Simon ES, Doniger GM, Cummings JL. Practicality of a computerized system for cognitive assessment in the elderly. *Alzheimers Dement*. 2008;4(1):14-21. doi:10.1016/j.jalz.2007.09.008

49. Stumps A, Saad E, Rothlein D, Verfaellie M, DeGutis J. Characterizing developmental prosopagnosia beyond face perception: Impaired recollection but intact familiarity recognition. *Cortex*. 2020;130:64-77. doi:10.1016/j.cortex.2020.04.016

Appendix

Table 1. Search strategy in databases

Database	Search strategy	Filters
PubMed	((‘Famous faces test’ OR ‘Famous face*’ OR ‘Famous name*’ OR (famous faces AND (celebrities OR photograph*))) AND (test OR recall OR recognition OR retrieval OR tip-of-the-tongue OR ‘temporal gradient’ OR ribot OR remote OR ‘proper name*’)) AND (‘older adults’ OR elderly OR ‘healthy elderly’ OR ‘normal aging’ OR geriatric* OR ‘preclinical dementia’ OR ‘preclinical Alzheimer*’ OR ‘prodromal dementia’ OR ‘prodromal Alzheimer*’ OR ‘mild cognitive impairment’ OR ‘early dementia’ OR ‘early Alzheimer*’ OR Dementia OR ‘senile dementia’ OR ‘incident dementia’ OR ‘Alzheimer’s Disease’ OR ‘Alzheimer's Dementia’)	Language: English Publication date: last five years
PsychINFO	((‘Famous faces test’ OR ‘Famous face*’ OR ‘Famous name*’ OR (famous faces AND (celebrities OR photograph*))) AND (test OR recall OR recognition OR retrieval OR tip-of-the-tongue OR ‘temporal gradient’ OR ribot OR remote OR ‘proper name*’)) AND (‘older adults’ OR elderly OR ‘healthy elderly’ OR ‘normal aging’ OR geriatric* OR ‘preclinical dementia’ OR ‘preclinical Alzheimer*’ OR ‘prodromal dementia’ OR ‘prodromal Alzheimer*’ OR ‘mild cognitive impairment’ OR ‘early dementia’ OR ‘early Alzheimer*’ OR Dementia OR ‘senile dementia’ OR ‘incident dementia’ OR ‘Alzheimer’s Disease’ OR ‘Alzheimer's Dementia’)	Language: English Publication date: April 2018 - April 2023 *Additional:* Source type: Academic journals and dissertations (not: books)
Web of Science	ALL=(((‘Famous faces test’ OR ‘Famous face*’ OR ‘Famous name*’ OR (famous faces AND (celebrities OR photograph*))) AND (test OR recall OR recognition OR retrieval OR tip-of-the-tongue OR ‘temporal gradient’ OR ribot OR remote OR ‘proper name*’))) AND ALL=((‘older adults’ OR elderly OR ‘healthy elderly’ OR ‘normal aging’ OR geriatric* OR ‘preclinical dementia’ OR ‘preclinical Alzheimer*’ OR ‘prodromal dementia’ OR ‘prodromal Alzheimer*’ OR ‘mild cognitive impairment’ OR ‘early dementia’ OR ‘early Alzheimer*’ OR Dementia OR ‘senile dementia’ OR ‘incident dementia’ OR ‘Alzheimer’s Disease’ OR ‘Alzheimer's Dementia’))	Language: English Publication date: last five years

Embase	('famous faces test' OR 'famous face*' OR 'famous name*' OR (famous AND faces AND ('celebrities'/exp OR celebrities OR photograph*))) AND ('test'/exp OR test OR 'recall'/exp OR recall OR 'recognition'/exp OR recognition OR 'retrieval'/exp OR retrieval OR 'tip of the tongue' OR 'temporal gradient' OR ribot OR remote OR 'proper name*') AND ('older adults'/exp OR 'older adults' OR 'elderly'/exp OR elderly OR 'healthy elderly' OR 'normal aging' OR geriatric* OR 'preclinical dementia' OR 'preclinical alzheimer*' OR 'prodromal dementia' OR 'prodromal alzheimer*' OR 'mild cognitive impairment'/exp OR 'mild cognitive impairment' OR 'early dementia' OR 'early alzheimer*' OR 'dementia'/exp OR dementia OR 'senile dementia'/exp OR 'senile dementia' OR 'incident dementia'/exp OR 'incident dementia' OR 'alzheimer disease'/exp OR 'alzheimer disease' OR 'alzheimer dementia'/exp OR 'alzheimer dementia')	Language: English Publication date: 2018 – 2023* *Additional:* Publication type: Article + Review (not: Conference Abstract or Paper)

[a] For the Embase search the term ‘’Alzheimer Disease’’ was used, as proposed by the database itself, since the term ‘’Alzheimer’s Disease’’ was not valid. The same holds for ‘’Alzheimer’s Dementia’’ which was replaced by ‘’Alzheimer Dementia’’.

[b] Embase had no option to further specify the date, therefore articles with a publication date >5 years were manually de-selected.