Stroke

CLINICAL AND POPULATION SCIENCES

New Picture Stimuli for the NIH Stroke Scale: A Validation Study

Melissa D. Stockbridge, PhD, CCC-SLP; Lindsey Kelly, MA, CCC-SLP; Sarah Newman-Norlund, MA; Brian White, AA; Marianne Bourgeois, BSN; Elizabeth Rothermel, BS; Julius Fridriksson, PhD, CCC-SLP; Patrick D. Lyden, MD; Argye E. Hillis, MD, MA

BACKGROUND: The National Institutes of Health Stroke Scale is a widely accepted tool for structured graded neurological examination of stroke or suspected stroke in the hyperacute setting. Concerns have arisen about the use of its picture stimuli in a contemporary and global health context. Here, we present new stimuli prepared to serve the needs of stroke providers worldwide: the precarious painter image description and updated objects for naming.

METHODS: This was a validation study of 101 healthy fluent English speakers. Participants were reached by the Johns Hopkins Outpatient Center, the University of South Carolina, and Prisma Health from 2022 to 2023 and included residents of the United States, Germany, Canada, the United Kingdom, Australia, and Zambia. Participants were recorded in person or via video conferencing when asked to describe the new picture, while a subset named seven illustrations. Multivariate analyses of variance were used for primary analyses. In a complementary investigation, 299 attendees of the 2023 International Stroke Conference were asked about their preference for the existing or new stimuli and why.

RESULTS: Each of the 44 content units from the picture description was included by at least 5% of respondents in the demographically representative subsample. Performance was similar across healthy participants irrespective of age, sex, race, ethnicity, or education. Typical descriptions were characterized by an average of 23 content units (SD=5) conveyed with 167 syllables (SD=79). The new naming stimuli were recognized by 100% of participants from many countries as being familiar and identifiable, and names provided in response to the task were highly convergent. The majority of stroke health care providers preferred both the precarious painter and naming stimuli.

CONCLUSIONS: The description of the new National Institutes of Health Stroke Scale picture, the precarious painter, results in rich samples among healthy speakers that will provide an appropriate basis for the detection of language deficits.

GRAPHIC ABSTRACT: A graphic abstract is available for this article.

Key Words: adult ■ aphasia ■ communication ■ language ■ language tests ■ surveys and questionnaires

In the more than 30 years since its debut, the National Institutes of Health Stroke Scale (NIHSS)[1,2] has grown to near universal use for structuring graded neurological examinations of stroke or suspected stroke in the hyperacute setting and become the gold standard for appraising poststroke function and change in neurological function.[3] It has demonstrated validity[4] and low-to-good reliability among providers,[5–7] which improves with training.[2,8] Importantly for research, acute NIHSS scores significantly predict both short- and long-term outcomes.[9,10]

See related article, p 452

Beyond institutionally addressable issues related to personnel training, prior authors have identified other areas of concern within the NIHSS rubric and materials. The score is biased by stroke lateralization. That is, individuals with left-hemisphere strokes score more highly than those with right-hemisphere strokes, even when controlling for both age and overall stroke volume.[11–15] This is

Correspondence to: Melissa D. Stockbridge, PhD, CCC-SLP, Department of Neurology, Johns Hopkins University School of Medicine, 600 N Wolfe St Phipps 446, Baltimore, MD 21287. Email md.stockbridge@jhmi.edu

This manuscript was sent to Emmanuel Touzé, Guest Editor, for review by expert referees, editorial decision, and final disposition.
Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/STROKEAHA.123.044384.
For Sources of Funding and Disclosures, see page 450.

Stroke is available at www.ahajournals.org/journal/str

Nonstandard Abbreviations and Acronyms

CU	content unit
NIHSS	National Institutes of Health Stroke Scale

due to the potential of the test to capture the effects of language deficits across multiple items, whereas deficits of communication and cognition more commonly associated with right-hemisphere damage are assessed only in the item neglect/inattention (neglect occurs in ≈38% of patients after right-hemisphere stroke versus 18% after left-hemisphere stroke).[16]

To evaluate the best language of a patient, administration of the NIHSS includes asking the patient to describe the cookie theft picture, originally a stimulus in the Boston Diagnostic Aphasia Evaluation,[17] to read words and sentences written on a card, and to identify the names of items (cacti, a glove, a chair, a key, a hammock, and a feather). The picture description has unique value because easily scored measures of content (content units [CUs]; concepts mentioned by healthy controls describing this picture),[18] efficiency with which patients convey key content (syllables/CUs), and the ratio of left:right CUs (to assess left or right neglect) are strongly correlated with lesion volume in both left- and right-hemisphere strokes.[12]

As the use of the NIHSS has expanded into communities of speakers of Chinese, German, Spanish, Cantonese, Estonian, Hindi, Hungarian, Italian, Marathi, Portuguese, Korean, Kannada, and Telugu (among others), the acceptability of these stimuli both socially and culturally also has been an area of scrutiny.[3,19] After all, not all countries have cacti. Moreover, describing a picture reminiscent of suburban life in midcentury America, wherein a dress-clad woman is gazing wistfully at her lawn from her kitchen window while her sink overflows and her children conspire furtively to access top-shelf cookies, does not similarly inspire people whose lived experiences may bear little resemblance to that depiction. Thus, it is timely, or perhaps overdue, that the NIHSS adopts a more contemporary and inclusive set of stimuli, easily recognized worldwide, to support more consistent interpretation globally. Prior groups have developed alternative pictures or other stimuli for the purposes of eliciting descriptive language,[20–30] but these have been developed especially for use in specific populations.

In a multidisciplinary and internationally representative initiative by the National Institute of Neurological Disorders and Stroke–supported investigators, a new picture has been commissioned from Elizabeth "Liddy" Rothermel of Apex Innovations, LLC: the precarious painter (Figure 1 (left); see Supplemental Material for high-resolution portable document format file suitable for printing). The illustration, which takes advantage of a full modern range of grayscale tones to increase visual interest, depicts a young person falling from a stepladder while painting a wall. The overall goal for creating the new picture was to depict a universally understood and relatable scene, with both major and minor areas of interest for the patient to comment on. The responses of a patient to these areas of interest, placed throughout each quadrant of the image (top/bottom, left/right), were intended to elicit further insight into relevant and affected areas of the brain. In

Figure 1. The precarious painter and naming items.

Stroke. 2024;55:443–451. DOI: 10.1161/STROKEAHA.123.044384

one unifying interpretation of the events, a car accident visible outside of the window has awoken a dog that has begun chasing a mouse through the room, inadvertently knocking over a bucket of paint and the ladder along the way. As a spiritual successor of the cookie theft image, the main action in the illustration is enriched by multiple smaller examples of disarray (eg, letters and pawprints left on the floor and a crooked framed picture of a cookie jar) and the illustration is, as a whole, open to differing interpretations of causation by the viewer.

Additionally, new object pictures were drawn to represent items that are recognized worldwide to replace the current object pictures (also originally from the Boston Diagnostic Aphasia Examination). We originally included seven items: mouse, cloud, bucket, leaf, traffic light, bicycle, and bridge (Figure 1 [right]).

Our purpose here is to evaluate the new stimuli for use in the NIHSS worldwide. Our aims were (1) to establish a rubric for assessing descriptions of the new picture, (2) to examine healthy performance on picture description and picture naming of the stimuli, and (3) to determine whether providers preferred the new stimuli to the existing stimuli.

METHODS

Anonymized data are available upon request to the authors, subject to review by the Johns Hopkins University School of Medicine institutional review board resulting in a formal data-sharing agreement. The descriptions of the precarious painter illustration were elicited from a large convenience sample of neurologically healthy participants. The first step in the analysis was to establish an inventory of the content most frequently identified across picture descriptions provided by participants (CUs; phase I). This was done using a demographically representative subsample of the total participants. Second, the rubric derived from the inventory was used to provide a quantifiable score of the information present in descriptions produced by the full sample of participants (phase II). Additionally, we evaluated name agreement in a subset of the same participants and recognition of the object pictures by individuals from many different countries. Finally, we surveyed an opportunity sample of attendees of the International Stroke Conference in 2023 through a Web-based survey to determine whether stroke providers (eg, physicians, nurses, stroke coordinators) preferred the new stimuli or the previous stimuli. Data are described in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology initiative.[31]

Recruitment of a Healthy Representative Sample

Sample picture descriptions were elicited from healthy, fluent English speakers. Participants were a convenience sample recruited from the community at large and from family members accompanying patients to visits at the Johns Hopkins Outpatient Center, the University of South Carolina, and Prisma Health from 2022 to 2023. Healthy volunteers recruited outside of the immediate geographic area were reached either using video conferencing tools (eg, Zoom, FaceTime) or through local affiliates of the Johns Hopkins Stroke Cognitive Outcomes and Recovery Laboratory, who administered the task in person. Exclusion criteria included previous neurological disease or diagnoses (eg, stroke, moderate-severe brain injury, Parkinson disease, dementia), blindness or uncorrected vision loss, deafness or hard of hearing individuals, or any other self-reported condition or medication thought to impact cognition or language, which were excluded (no individual who provided a sample was later excluded). Self-described demographic information was collected, including age (in years), sex, race, ethnicity (Hispanic/non-Hispanic), and educational attainment. The study was approved by the Johns Hopkins University institutional review board, which determined that written consent for participation was not required.

One hundred one samples were collected by researchers based at Johns Hopkins University and the University of South Carolina. Analyses were conducted in 2 phases. First, a subsample was identified and used to establish the CUs in the picture (N=50). Then, the full sample of descriptions of participants (both those in the subsample and additional participants who were recruited) was analyzed using the CUs identified in the first phase.

Members of the phase I subsample were selected to reflect the age distribution of stroke[32] and sex, race, ethnicity, and education distribution reflecting US census data from 2020 to 2023. Although other racial categories are included in the census data: ≥2 races, Native American or Alaskan Natives, Native Hawaiian or other Pacific Islander, or some other race not specified, none of the 101 individuals approached self-described in these ways in an open-ended prompt. Although the majority of participants were from the United States (rural and urban regions of east, west, north, south, and central states), we also obtained descriptions by English-speaking people from Germany, Canada, the United Kingdom, Australia, and Zambia. Target numbers of individuals were calculated out of 50 individuals based on the census percentages. Target and achieved representations of each demographic variable within the subsample are summarized in Table 1. Compared with the sample used in phase I, those in phase II had a lower mean age of 54 years (±18) and a higher mean number of years of education of 16 (±3).

Precarious Painter Validation

Acquisition of Language Samples and Establishment of CUs

Participants were presented with the precarious painter drawing and asked to "describe everything that they saw going on in the picture," consistent with the current guidelines for administering the NIHSS language assessment for item 11. Participants were given 2 minutes to respond, and all responses were audio-recorded for later transcription and analysis.

The process of identifying CUs was consistent with the process described in the literature when establishing visual stimuli used to elicit discourse samples.[20,33] Forty-four CUs were identified based on a visual examination of the image: 17 on the right side of the picture, 22 on the left side of the picture, and 5 that were present on both sides or reflected general appraisal of the scene in the image (eg, the scene is a mess

Table 1. Demographics of Representative Subsample Used to Establish Content Units

	Target N (%)	Sample N (%)	Sample mean (SD)
Age, y			
<65	19 (38%)	20 (40%)	62 (18)
≥65	31 (62%)	30 (60%)	
Sex			
Female	26 (51%)	26 (52%)	
Male	24 (49%)	24 (48)	
Race			
White	31 (62%)	38 (76%)	
Black	6 (12%)	8 (16%)	
Asian	3 (6%)	4 (8%)	
Other	10 (20%)	0 (0%)	
Ethnicity			
Hispanic	9 (19%)	3 (6%)	
Non-Hispanic	41 (81%)	46 (92%)	
Education			
HS or less	18 (37%)	17 (34%)	14 (3)
Some college	13 (25%)	13 (26%)	
College graduate	12 (24%)	10 (20%)	
Post-graduate	7 (14%)	10 (20%)	

A small number of demographic characteristics were not disclosed by participants. These individuals are excluded from the analysis of that dimension. Missing ethnicity for 1 individual. HS indicates high school.

or catastrophe). CUs included nouns (eg, dog, door), verbs (eg, chase, crash), modifiers (eg, unstable, crooked/askew), or prepositional phrases (eg, in the corner, on the floor).

The first version of the illustration included a car hitting a fire hydrant, but this was modified early in data collection to be a car hitting a tree, as participants from other countries noted that fire hydrants are not internationally ubiquitous or recognized. The CU for the "fire hydrant" and "tree" were collapsed into a single target for the purposes of both establishing the CUs of the picture and scoring the full series of samples. However, future samples should be considered using only "tree." CUs mentioned by <5% of sample individuals were discarded.

Language Sample Analysis

Utilizing the 44 CUs established in the first phase, the remaining samples were transcribed, and all 101 picture descriptions were analyzed to examine the 4 key discourse variables: total CUs, left:right proportion of CUs, syllables, and syllables per CU, a measure of discourse informational efficiency.[18,20,34,35] The 2 certified speech-language pathologists, M.D.S. and L.K., showed 97.7% point-to-point agreement in scoring CUs in the descriptions of ≈10% of the samples (11 participants who utilized 484 total CUs). As the total number of CUs on the left and right differed, counts were first divided by the total on each side, and then, the ratio was calculated.

To examine the relationships between the 4 related outcome variables and 5 individual factors of interest (age, sex, race, ethnicity, and education), multivariate analyses of variance were calculated for each of the 5 factors, considering the 4 outcome variables together. Multivariate analysis of variance is used for comparing multivariate means (>1 dependent variable). Pillai trace is measured on a scale of 0 to 1, with higher values constituting stronger evidence of effect. Each multivariate analysis of variance was considered at an α=0.05/5=0.01 level of significance, 2-tailed. Then, between-subject effects were examined using ANOVA. ANOVA is used to determine the relationship between >2 factors and a single dependent variable. Each ANOVA was considered at an α=0.05/(4*5)=0.002 level of significance, 2-tailed (Bonferroni correction for 5 factor and 4 outcome variables). Given the conservative criteria, trends in the data that did not reach significance also are discussed.

Object Picture Evaluation

A subset of participants from the above sample (n=41) were asked to name the new objects to evaluate name agreement. Individuals asked to complete this additional task were selected at random among those who agreed to participate at Johns Hopkins University, and no one who was asked declined to complete the task. The object pictures also were shown to 50 participants at 2 international conferences (the Academy of Aphasia and the European Stroke Organization Congress) in 2023 simply to determine whether people from various countries would recognize the objects. Participants from the United Kingdom, Germany, Sweden, Turkey, Australia, Canada, Colombia, Brazil, Chili, China, Japan, Russia, Nigeria, and Zambia were shown the pictures and asked to check the ones that they recognized (and would be able to name in any language).

Evaluation of Stroke Provider Preference

We surveyed attendees of the International Stroke Conference in 2023 (in Dallas) through a Web-based survey to determine whether stroke providers (physicians, nurses, and therapists) preferred the precarious painter picture or the cookie theft picture and whether they preferred the new set of object pictures or the previous set. They were asked to provide their discipline and their preference and check an answer (among 4) for why they preferred the original or the new stimuli.

RESULTS

Precarious Painter Validation

Phase I: Establishing CUs Using a Demographically Representative Sample

All 44 CUs were included by at least 5% of samples within the 50-person representative subsample (Figure 2). All CUs were retained for analysis within the full sample. A list of CUs usable as a rubric for future administrations and an Excel file that automatically calculates totals and the ratio of left:right proportions are available in the Supplemental Material.

Performance was consistent among healthy English speakers irrespective of demographic differences (Table 2). There were no significant differences as a function of age: Pillai trace, 0.04, F(4, 45), 0.44, and *P*=0.78; sex: Pillai trace, 0.09, F(4, 45), 1.12, and *P*=0.36; race: Pillai trace, 0.27, F(8, 90), 1.78, and *P*=0.09; ethnicity:

	Stroke. 2024;55:443–451. DOI: 10.1161/STROKEAHA.123.044384

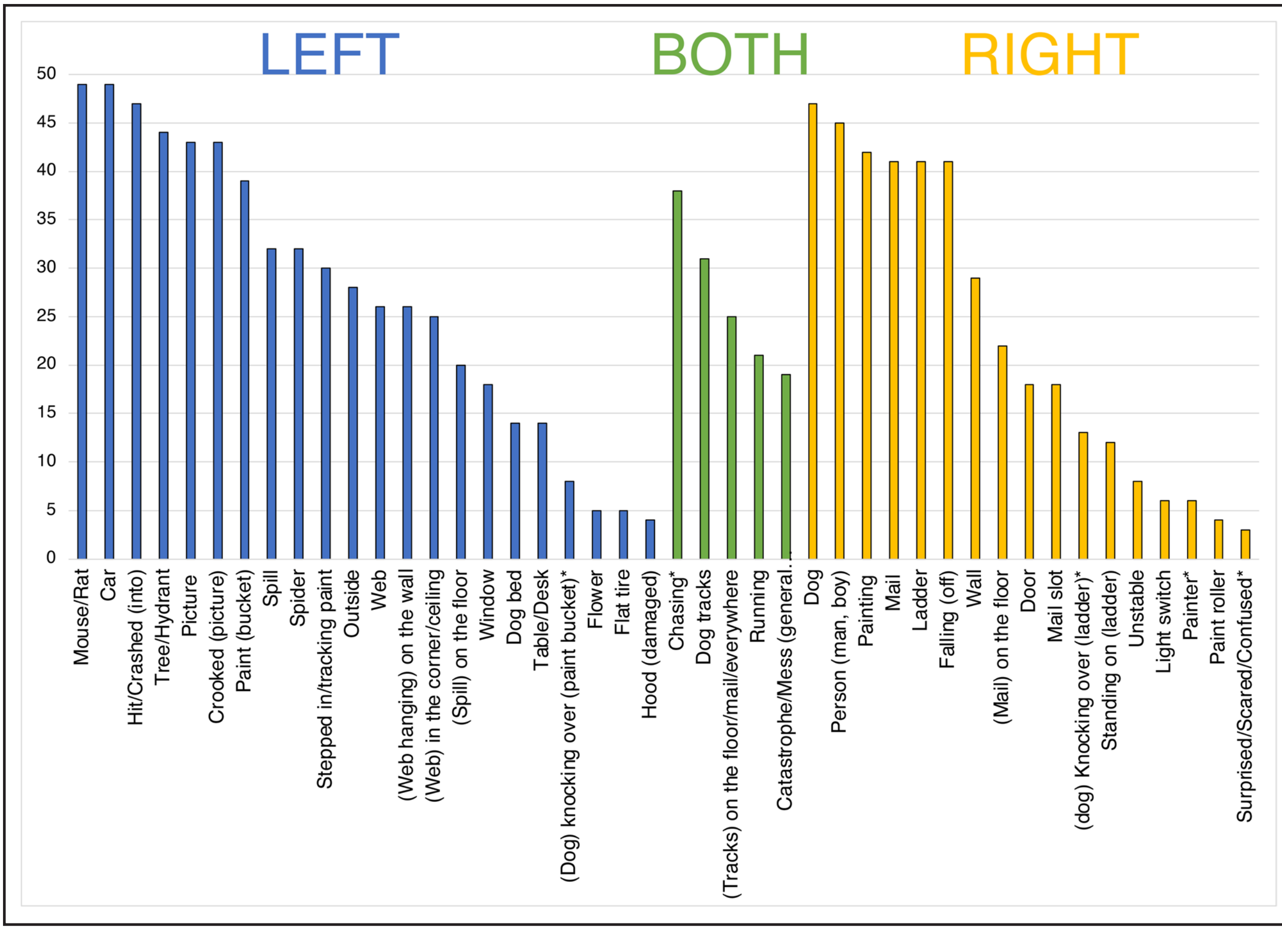

Figure 2. Number of samples that included each content unit.

Pillai trace, 0.11, F(8, 90), 0.64, and P=0.74; or education: Pillai trace, 0.19, F(12, 135), 0.77, and P=0.68.

Exploratory univariate analyses of variance conducted to contextualize the patterns of performance observed in the sample data also found no significant differences as a function of age, sex, ethnicity, or educational attainment for any of the 4 discourse variables. However, a trend was noted in the differences by race in total CUs, F(2, 47), 3.28, P=0.05, and η_P^2=0.12, and proportion of left:right CUs, F(2, 47), 3.8, P=0.03, and η_P^2=0.14, driven by fewer CUs and a higher proportion of left-to-right content described among Black responders than Asian or White responders.

Phase II: Precarious Painter Performance Within the Full Sample

Using the full sample of participants, no significant differences were observed as a function of age: Pillai trace, 0.01, F(4, 96), 0.32, and P=0.87; ethnicity: Pillai trace, 0.02, F(4, 95), 0.43, and P=0.79; or education, Pillai Trace, 0.11, F(12, 285), 0.87, and P=0.58. There was a trend toward an effect of sex, Pillai trace, 0.11, F(4, 95), 2.97, P=0.02, and η_P^2=0.11, and a significant effect of race, Pillai trace, 0.20, F(8188), 2.66, P=0.009, and η_P^2=0.10.

Table 2. Performance Within Demographically Representative Sample

	N	CU	Left:right	Syllables	Syllables:CU
Age, y					
<65	20	22 (6)	1.2 (0.4)	162 (84)	7.6 (3.4)
≥65	30	23 (5)	1.3 (0.4)	170 (77)	7.2 (2.7)
Sex					
Female	26	23 (5)	1.2 (0.4)	173 (88)	7.4 (2.7)
Male	24	23 (6)	1.3 (0.4)	161 (69)	7.4 (3.3)
Race					
White	38	23 (5)	1.2 (0.3)	157 (84)	7.0 (2.7)
Black	8	18 (7)	1.6 (0.5)	140 (55)	8.1 (4.2)
Asian	4	24 (3)	1.2 (0.4)	218 (37)	9.3 (2.5)
Ethnicity					
Hispanic	3	21 (2)	1.0 (0.1)	148 (71)	7.0 (2.8)
Non-Hispanic	46	23 (5)	1.3 (0.4)	168 (81)	7.4 (3.0)
No response	1	30	1.2	192	6.4
Education					
HS or less	17	21 (6)	1.4 (0.5)	143 (69)	7.0 (3.4)
Some college	13	24 (5)	1.2 (0.3)	179 (96)	7.3 (2.6)
College graduate	10	23 (6)	1.2 (0.5)	155 (44)	7.1 (1.8)
Post-graduate	10	25 (5)	1.1 (0.2)	204 (90)	8.4 (3.8)
Total	50	23 (5)	1.2 (0.4)	167 (79)	7.4 (3.0)

CU indicates content unit; and HS, high school.

Table 3. Precarious Painter Performance

	N	CU	Left:right	Syllables	Syllables:CU
Age, y					
<65	67	23 (5)	1.2 (0.4)	163 (77)	7.1 (2.7)
≥65	34	23 (5)	1.3 (0.4)	171 (76)	7.3 (2.7)
Sex					
Female	64	24 (5)	1.2 (0.3)	169 (83)	7.0 (2.5)
Male	36	22 (6)	1.3 (0.4)	161 (63)	7.6 (3.0)
Race					
White	81	23 (5)	1.2 (0.3)	167 (80)	7.1 (2.5)
Black	12	21 (7)	1.6 (0.5)	149 (51)	7.7 (3.4)
Asian	6	22 (6)	1.1 (0.4)	177 (72)	7.9 (2.9)
Ethnicity					
Hispanic	3	21 (2)	1.0 (0.1)	148 (71)	7.0 (2.8)
Non-Hispanic	97	23 (5)	1.2 (0.4)	166 (77)	7.2 (2.7)
Education					
HS or less	18	21 (6)	1.4 (0.5)	141 (68)	7.0 (3.3)
Some college	15	24 (5)	1.2 (0.3)	170 (93)	6.9 (2.6)
College graduate	30	23 (6)	1.2 (0.4)	162 (63)	7.1 (1.8)
Post-graduate	37	24 (5)	1.2 (0.4)	178 (84)	7.3 (3.1)
Total	101	23 (5)	1.2 (0.4)	166 (76)	7.2 (2.7)

A small number of demographic characteristics were not disclosed by participants. These individuals are excluded from the analysis of that dimension. Missing sex, ethnicity, and education each for 1 individual. Missing race for 2 individuals. CU indicates content unit; and HS, high school.

Exploratory univariate ANOVAs were conducted to contextualize the patterns of performance observed, and the central tendency for each variable is presented in Table 3. No differences were noted in performance as a function of age, ethnicity, or educational attainment for any of the 4 discourse variables. A trend was observed in the proportion of left:right CUs by sex: $F(1, 98)$, 5.6, P=0.02, and η_P^2=0.05. No other differences in performance by sex approached significance. Sex differences were driven by a slightly higher proportion of left-to-right content among men than women. There was a significant effect of race on the proportion of left:right CUs: $F(2, 96)$, 9.61, P<0.001, and η_P^2=0.17. Race differences were driven by a higher proportion of left-to-right content described among Black responders than Asian or White responders. Despite doubling the individuals included in the sample between phases I and II, the average performance and SD were nearly identical across all 4 discourse variables.

Object Picture Validation

There was a high name agreement for 6 of the new pictures (Table 4). All participants named bridge, cloud, leaf, and bicycle with a response that included the anticipated target name. The mouse drawing was named mouse or rat, and the bucket was named bucket or pail by all participants, which were objectively equally acceptable. There are known regional differences in how these words are used, particularly bucket and pail. However,

Table 4. Picture Naming Convergence

	N	%
Bucket/pail		
Bucket	29	71
Pail	7	17
Both bucket and pail provided	5	12
Bicycle/bike	41	100
Cloud(s)		
Cloud	21	51
Clouds	20	49
Traffic light/traffic signal/stoplight		
Traffic light(s)	20	49
Traffic signal	7	17
Stoplight	5	12
Car signal	1	2
Red light, street lamp	1	2
Robot	1	2
Speaker	1	2
Stoplight sign	1	2
Stop signal	1	2
Stoplight, traffic light	1	2
Streetlight	1	2
Traffic control device	1	2
Leaf		
Leaf	39	95
Feather, leaf	1	2
Not provided	1	2
Mouse/rat		
Mouse	30	73
Mice	2	5
Rat	6	15
Both mouse and rat provided	3	7
Bridge		
Bridge	32	78
Footbridge	3	7
Little bridge	2	5
Walking bridge	2	5
Small bridge	1	2
Wooden bridge	1	2

a variety of names (all correct in the countries of the participants) were provided for traffic light, for example, traffic light, stoplight, traffic signal, streetlight, robot, and traffic robot.

Provider Preference

Among attendees at the International Stroke Conference 2023, 299 people responded to the Web-based survey. The precarious painter was preferred over the cookie theft picture by 60% of respondents. The new picture was preferred across all specialties, including hospital stroke coordinators (64%), patient educators (62%), nurses (61%), and neurologists (57%). For responders who indicated reasons for the preference, the main reasons were "more items to identify" (n=35), "more updated modern/contemporary" (n=33), and "more inclusive/less biased/stereotypical" (n=40). Likewise, the new object pictures were preferred across all specialties, including hospital stroke coordinators (91%), patient educators (100%), nurses (91%), and neurologists (79%). Reasons for the preference were "easier/more common to identify" (n=56), "simpler/clearer drawings" (n=25), "more updated/modern/contemporary" (n=21), and "more inclusive culturally" (n=17).

DISCUSSION

The purpose of this investigation was to introduce new stimuli developed for international use in the NIHSS and provide confirmation that diverse, neurologically unimpaired, English-speaking adults similar to those who experience strokes perform similarly to one another. For the precarious painter, each of the 44 CUs was included by at least 5% of respondents in the demographically representative subsample. As anticipated, performance was similar across the 50-person subsample and the 101-person full sample, supporting its reliability for language assessment. Although there are challenges to directly contrasting language samples across contexts, adults in the present sample produced more content in response to the precarious painter (23 CUs) than previously described healthy adults describing the cookie theft (14.7–18 CUs).[18] A curious feature of these data was the emergence of a small but significant difference in the proportion of left:right CUs by race in post hoc ANOVA (η_p^2=0.17). As these individuals are unimpaired, there is no reasonable expectation that this was the result of any underlying clinical but unidentified difference (ie, undiagnosed neglect). It is entirely possible that this effect was spurious given the small sample size and the relatively large number of statistical comparisons made. However, a visual inspection of the data suggested that the pattern observed by race was the result of Black individuals who included fewer items on the right than other groups (as opposed to highly detailed descriptions of the left). If it is a true difference, it may be due to the interaction between the culture of the viewer and the image contents.[36,37] Differences in the perceived importance of events can be cultural and influence what an individual determines is most salient. An imperative ongoing and future direction is to understand how individual factors interact in formal language assessment.

The new object pictures were recognized by 100% of participants from many countries (across all major continents) as being familiar and namable. Name agreement in English was functionally unanimous for a single name for 4 of the items. For 2 items, 1 of 2 acceptable names was given reliably in English. However, given the variety of names for traffic lights, we propose to drop this item from the stimuli.

The majority of stroke providers preferred the new stimuli. Nurses, stroke coordinators, and stroke educators were more in favor of the precarious painter than stroke physicians. Even among physicians, however, a majority preferred precarious painter over the cookie theft picture. Common reasons for their preference included more items to identify, more contemporary, and more inclusive/less stereotypical. A vast majority of all providers preferred the new object pictures.

Limitations of the present study include that most participants were from the United States. We did include volunteers from many other countries, both in providing language responses and provider preferences, because the NIHSS is used globally. However, providers in each country will need to identify responses that are considered accurate versus inaccurate in their country or language. We excluded 1 item for which an accurate response (in some English-speaking countries), a "robot," might be considered a reflection of aphasia in other English-speaking countries. Sentence stimuli for reading will also need to be translated into local languages. Another limitation is the small sample size here; further validation studies are warranted.

Assessment stimuli should be revised when a sufficient shift has rendered the target no longer widely recognized or culturally neutral. Recent examples include the replacement of the calculator in a test of preschool recognition[38] and the noose in the Boston Naming Test.[39] These examples suggest stimuli benefit from critical consideration with every generation (20–30 years). Replacing the cookie theft, which was introduced in 1972, with a new image in the highly visible context of the NIHSS will lend further momentum to calls for the retirement of the cookie theft picture. Evaluating the use of the precarious painter in clinical populations is the next step in establishing its validity and utility. Data collection in acute, subacute, and chronic stroke is ongoing. We will determine if the numbers of CU, syllables/CU, and right:left CU correlate with lesion volume, as we found for the cookie theft picture.[12] If so, a checklist of CU can be used (taking <2 minutes in previous picture description tasks) to provide a quick assessment of severity that complements the NIHSS total score in both left- and right-hemisphere strokes. We hope that the new stimuli will serve the next generations of clinicians and researchers utilizing the NIHSS as effectively as the previous stimuli have for the last 50 years.

ARTICLE INFORMATION

Received June 30, 2023; final revision received November 14, 2023; accepted November 29, 2023.

Affiliations

Department of Neurology (M.D.S., L.K., A.E.H.) and Department of Physical Medicine and Rehabilitation (A.E.H.), Johns Hopkins University School of Medicine, Baltimore, MD. Department of Communication Sciences and Disorders, Arnold School of Public Health, University of South Carolina, Columbia (S.N.-N., J.F.). Apex Innovations LLC, Lafayette, LA (B.W., M.B., E.R.). Department of Physiology and Neuroscience, Zilkha Neurogenetic Institute, Keck School of Medicine of University of Southern California, Los Angeles (P.D.L.). Department of Neurology, Keck School of Medicine, Los Angeles, CA (P.D.L.). Department of Cognitive Science, Krieger School of Arts and Sciences, Johns Hopkins University, Baltimore, MD (A.E.H.).

Acknowledgments

The authors acknowledge the advice received from Jamila Minga on the interpretation of our findings in light of the Hall theory of high and low context cultures.

Sources of Funding

This work was supported by the National Institutes of Health/National Institute on Deafness and Other Communication Disorders grants P50 DC014664, R01 DC05375, R01 DC015466, and R01 DC011739, U24 NS130600, and UG3 NS119199 (to Dr Lyden). Additionally, this work was supported by the University of South Carolina Excellence Initiative for Aging Brain Cohort at University of South Carolina.

Disclosures

Apex Innovations is an accredited provider of educational services for health care facilities, which provides free training videos developed by neurologists from the University of Cincinnati, the University of Iowa, and the National Institute of Neurological Disorders and Stroke (NINDS) in the reliable and valid administration of the National Institutes of Health (NIH) Stroke Scale. The content of the videos is based on medical science, accepted standards of care, and best practice guidelines. Neither they nor the authors receive any commercial support associated with these educational activities or the NIH Stroke Scale, which is made freely available by the NINDS. Dr Lyden is a paid consultant to Apex Innovations. Dr Hillis receives compensation from the American Heart Association as the editor-in-chief of *Stroke* and from Elsevier as an associate editor of *PracticeUpdate Neurology*. All authors receive salary support from NIH (National Institute on Deafness and Other Communication Disorders [NIDCD] and NINDS) through grants.

Supplemental Material

Precarious Painter High-Resolution Portable Document Format File
Naming Objects High-Resolution Portable Document Format File
Precarious Painter Microsoft Word Rubric for Scoring Content Units
Precarious Painter Microsoft Excel Rubric With Calculation of Content Units and Ratios

REFERENCES

1. Brott T, Adams HP Jr, Olinger CP, Marler JR, Barsan WG, Biller J, Spilker J, Holleran R, Eberle R, Hertzberg V. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke*. 1989;20:864–870. doi: 10.1161/01.str.20.7.864
2. Lyden P, Brott T, Tilley B, Welch K, Mascha E, Levine S, Haley E, Grotta J, Marler J. Improved reliability of the NIH stroke scale using video training NINDS TPA stroke study group. *Stroke*. 1994;25:2220–2226. doi: 10.1161/01.str.25.11.2220
3. Lyden P. Using the National Institutes of Health Stroke Scale: a cautionary tale. *Stroke*. 2017;48:513–519. doi: 10.1161/STROKEAHA.116.015434
4. Lyden P, Raman R, Liu L, Emr M, Warren M, Marler J. National Institutes of Health Stroke Scale certification is reliable across multiple venues. *Stroke*. 2009;40:2507–2511. doi: 10.1161/STROKEAHA.108.532069
5. Josephson SA, Hills NK, Johnston SC. NIH Stroke Scale reliability in ratings from a large sample of clinicians. *Cerebrovasc Dis*. 2006;22:389–395. doi: 10.1159/000094857
6. Goldstein LB, Samsa GP. Reliability of the National Institutes of Health Stroke Scale: extension to non-neurologists in the context of a clinical trial. *Stroke*. 1997;28:307–310. doi: 10.1161/01.str.28.2.307
7. McLoughlin A, Olive P, Lightbody CE. Reliability of the National Institutes of Health Stroke Scale. *Br J Neurosci Nurs*. 2022;18:S3–S10. doi: 10.12968/bjnn.2022.18.Sup5.S3
8. Schmülling S, Grond M, Rudolf J, Kiencke P. Training as a prerequisite for reliable use of NIH Stroke Scale. *Stroke*. 1998;29:1258–1259. doi: 10.1161/01.str.29.6.1258
9. Adams HP, Davis P, Leira E, Chang K-C, Bendixen B, Clarke W, Woolson R, Hansen MD. Baseline NIH Stroke Scale score strongly predicts outcome after stroke: a report of the trial of Org 10172 in Acute Stroke Treatment (TOAST). *Neurology*. 1999;53:126–126. doi: 10.1212/wnl.53.1.126
10. Schlegel D, Kolb SJ, Luciano JM, Tovar JM, Cucchiara BL, Liebeskind DS, Kasner SE. Utility of the NIH Stroke Scale as a predictor of hospital disposition. *Stroke*. 2003;34:134–137. doi: 10.1161/01.str.0000048217.44714.02
11. Vitti E, Kim G, Stockbridge MD, Hillis AE, Faria AV. Left hemisphere bias of NIH Stroke Scale is most severe for middle cerebral artery strokes. *Front Neurol*. 2022;13:912782. doi: 10.3389/fneur.2022.912782
12. Agis D, Goggins MB, Oishi K, Oishi K, Davis C, Wright A, Kim EH, Sebastian R, Tippett DC, Faria A, et al. Picturing the size and site of stroke with an expanded National Institutes of Health Stroke Scale. *Stroke*. 2016;47:1459–1465. doi: 10.1161/STROKEAHA.115.012324
13. Woo D, Broderick JP, Kothari RU, Lu M, Brott T, Lyden PD, Marler JR, Grotta JC. Does the National Institutes of Health Stroke Scale favor left hemisphere strokes? NINDS t-PA stroke study group. *Stroke*. 1999;30:2355–2359. doi: 10.1161/01.str.30.11.2355
14. Fink JN, Selim MH, Kumar S, Silver B, Linfante I, Caplan LR, Schlaug G. Is the association of National Institutes of Health Stroke Scale scores and acute magnetic resonance imaging stroke volume equal for patients with right- and left-hemisphere ischemic stroke? *Stroke*. 2002;33:954–958. doi: 10.1161/01.str.0000013069.24300.1d
15. Gottesman RF, Kleinman JT, Davis C, Heidler-Gary J, Newhart M, Hillis AE. The NIHSS-plus: improving cognitive assessment with the NIHSS. *Behav Neurol*. 2010;22:11–15. doi: 10.3233/BEN-2009-0259
16. Esposito E, Shekhtman G, Chen P. Prevalence of spatial neglect post-stroke: a systematic review. *Ann Phys Rehabil Med*. 2021;64:101459. doi: 10.1016/j.rehab.2020.10.010
17. Goodglass H, Kaplan E. *The Assessment of Aphasia and Related Disorders*. Lea & Febiger; 1972.
18. Yorkston KM, Beukelman DR. An analysis of connected speech samples of aphasic and normal speakers. *J Speech Hear Disord*. 1980;45:27–36. doi: 10.1044/jshd.4501.27
19. Steinberg A, Lyden PD, Davis AP. Bias in stroke evaluation: rethinking the cookie theft picture. *Stroke*. 2022;53:2123–2125. doi: 10.1161/STROKEAHA.121.038515
20. Berube S, Nonnemacher J, Demsky C, Glenn S, Saxena S, Wright A, Tippett DC, Hillis AE. Stealing cookies in the twenty-first century: measures of spoken narrative in healthy versus speakers with aphasia. *Am J Speech Lang Pathol*. 2019;28:321–329. doi: 10.1044/2018_AJSLP-17-0131
21. Lim SH, Guek TY, Woon FP, Tay DDY, Ho SS, Ng SC, De Silva DA. National Institutes of Health Stroke Scale: comparison of the original versus a modified version for Singapore culture. *Singapore Med J*. 2021;64:563. doi: 10.11622/smedj.2021126
22. Oh MS, Yu KH, Lee JH, Jung S, Ko IS, Shin JH, Cho SJ, Choi HC, Kim HH, Lee BC. Validity and reliability of a Korean version of the National Institutes of Health Stroke Scale. *J Clin Neurol*. 2012;8:177–183. doi: 10.3988/jcn.2012.8.3.177
23. Cheung RT, Lyden PD, Tsoi TH, Huang Y, Liu M, Hon SF, Raman R, Liu L. Production and validation of Putonghua- and Cantonese-Chinese language National Institutes of Health Stroke Scale training and certification videos. *Int J Stroke*. 2010;5:74–79. doi: 10.1111/j.1747-4949.2010.00411.x
24. Domínguez R, Vila JF, Augustovski F, Irazola V, Castillo PR, Rotta Escalante R, Brott TG, Meschia JF. Spanish cross-cultural adaptation and validation of the National Institutes of Health Stroke Scale. *Mayo Clin Proc*. 2006;81:476–480. doi: 10.4065/81.4.476
25. Cincura C, Pontes-Neto OM, Neville IS, Mendes HF, Menezes DF, Mariano DC, Pereira IF, Teixeira LA, Jesus PA, de Queiroz DC, et al. Validation of the National Institutes of Health Stroke Scale, modified Rankin scale and Barthel index in Brazil: the role of cultural adaptation and structured interviewing. *Cerebrovasc Dis*. 2009;27:119–122. doi: 10.1159/000177918
26. Sun TK, Chiu SC, Yeh SH, Chang KC. Assessing reliability and validity of the Chinese version of the stroke scale: scale development. *Int J Nurs Stud*. 2006;43:457–463. doi: 10.1016/j.ijnurstu.2005.07.004
27. Hussein HM, Abdel Moneim A, Emara T, Abd-Elhamid YA, Salem HH, Abd-Allah F, Farrag MA, Tork MA, Shalash AS, Ezz El Dein KH, et al. Arabic cross cultural adaptation and validation of the National Institutes of Health Stroke Scale. *J Neurol Sci*. 2015;357:152–156. doi: 10.1016/j.jns.2015.07.022
28. Prasad K, Dash D, Kumar A. Validation of the Hindi version of National Institutes of Health Stroke Scale. *Neurol India*. 2012;60:40–44. doi: 10.4103/0028-3886.93587
29. Pezzella FR, Picconi O, De Luca A, Lyden PD, Fiorelli M. Development of the Italian version of the National Institutes of Health Stroke Scale: It-NIHSS. *Stroke*. 2009;40:2557–2559. doi: 10.1161/STROKEAHA.108.534495

30. Burns M, Somers K, McElwaine P, Harbison J. National Institutes of Health Stroke Scale object naming test in a non-American English-speaking population. *Int J Stroke*. 2014;9:E35. doi: 10.1111/ijs.12348
31. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;61:344–349. doi: 10.1016/j.jclinepi.2007.11.008
32. Jackson G, Chari K. National Hospital Care Survey demonstration projects: Stroke inpatient hospitalizations. National Health Statistics Reports; no 132. Hyattsville, MD: National Center for Health Statistics. 2019.
33. Stockbridge MD, Newman R. Enduring cognitive and linguistic deficits in individuals with a history of concussion. *Am J Speech Lang Pathol*. 2019;28:1554–1570. doi: 10.1044/2019_AJSLP-18-0196
34. Keator LM, Faria AV, Kim KT, Saxena S, Wright AE, Sheppard SM, Breining BL, Goldberg E, Tippett DC, Meier E, et al. An efficient bedside measure yields prognostic implications for language recovery in acute stroke patients. *Cogn Behav Neurol*. 2020;33:192–200. doi: 10.1097/WNN.0000000000000238
35. Berube SK, Goldberg E, Sheppard SM, Durfee AZ, Ubellacker D, Walker A, Stein CM, Hillis AE. An analysis of right hemisphere stroke discourse in the modern cookie theft picture. *Am J Speech Lang Pathol*. 2022;31:2301–2312. doi: 10.1044/2022_AJSLP-21-00294
36. Hall ET. *Beyond Culture*. Anchor; 1976. doi: 10.1044/2022_AJSLP-21-00294
37. Hall ET, Hall MR. *Understanding Cultural Differences*. Intercultural Press; 1989.
38. Wiig EH, Secord W, Semel EM. *CELF Preschool 3: Clinical Evaluation of Language Fundamentals - Preschool*. In: Clinical Evaluation of Language Fundamentals - Preschool. Pearson, Inc; 2020.
39. Chokshi A, King JA. A-305 Boston naming test replacement item 48 (boomerang) in a clinical sample. *Arch Clin Neuropsychol*. 2022;37:1459–1459. doi: 10.1093/arclin/acac060.305