Digital Neuropsychological assessment of discourse production: an Interdisciplinary approach

Marco Begolo
marco.begolo@studenti.unipd.it
Department of Mathematics "Tullio Levi-Civita",
University of Padua
Padua, Italy

Ombretta Gaggi
gaggi@math.unipd.it
Department of Mathematics "Tullio Levi-Civita",
University of Padua
Human Inspired Technology-Research Centre, University of Padua
Padua, Italy

Sara Mondini
sara.mondini@unipd.it
Department of General Psychology, University of Padua
Human Inspired Technology-Research Centre, University of Padua

Sonia Montemurro
sonia.montemurro@phd.unipd.it
Department of General Psychology, University of Padua
Human Inspired Technology-Research Centre, University of Padua

ABSTRACT

In neuropsychology, the assessment of speech production has to be very accurate and precise when a patient shows disorganized discourse in conversation. Clinician's work in such cases, consists in a multi-tasking operation based on the observation and the quantification of a number of parameters of speech, occurring during the patient's discourse production. In this paper, we describe a m-health system for the administration of one of the tasks included in the Assessment of Pragmatic Abilities and Cognitive Substrates (APACS). It is a tablet application with a user-friendly interface, able to facilitate both the assessment and the scoring procedures. To our knowledge, this is the first study which proposes the Italian digital version of a tool for assessing different dimensions of pragmatic and paralinguistic features of discourse production.

CCS CONCEPTS

• **Applied computing → Health informatics**.

KEYWORDS

APACS test, digitalization

GoodTechs '19, September 25–27, 2019, Valencia, Spain

ACM ISBN 978-1-4503-6261-0/19/09…$15.00
https://doi.org/10.1145/3342428.3342656

ACM Reference Format:
Marco Begolo, Ombretta Gaggi, Sara Mondini, and Sonia Montemurro. 2019. Digital Neuropsychological assessment of discourse production: an Interdisciplinary approach. In *EAI International Conference on Smart Objects and Technologies for Social Good (GoodTechs '19), September 25–27, 2019, Valencia, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3342428.3342656

1 INTRODUCTION

The most recent literature on neuropsychological assessment demonstrates that tests need to be updated in order to improve efficiency and utility [8]. It involves the inclusion of new technologies, modern psychometric concepts and theoretical constructs.

Cognitive evaluation of speech production is very important when a neuropsychological patient shows disorganized discourse and disruptions in conversation due to neurological diseases. Clinician's work consists in a multi-tasking operation in which the full observation of the patient must be combined with the scoring of the patient's cognitive performance, considering different parameters of his/her speech.

Difficulties in discourse production have been found in many clinical populations, like in Alzheimer's Disease [1], Schizophrenia [4], Traumatic Brain Injury [2], Parkinson's Disease [17] and Multiple Sclerosis [6]. These individuals may show language difficulties in speaking, presented simultaneously within the patients' discourse, which makes the cognitive assessment highly demanding in a paper and pencil score-form. The purpose of the present work is the using of IT approaches in neuropsychology. This has a great potential, since it may optimize the accuracy and the detailed registration and scoring of the neuropsychological assessment.

In addition to language problems, people with neurological diseases often show low attentional resources compared to the normal population, which indicates the urgency to have the best cognitive assessment in a very short time in order to avoid errors due to patients' tiredness. In other words, the traditional long paper-pencil procedure needs to be improved with new forms. The computerization of cognitive tasks may optimize the time needed for:

- assessing different features of cognitive performance shown at the same time,
- assigning the specific score and
- calculating global scores.

The digitalization of cognitive tasks can definitely improve the patient observation, without substituting the clinician's observation, which remains essential for a reliable neuropsychological diagnosis. Computer-based instruments in neuropsychology, instead, can play a very important role for improving the technical aspects of assessment through user-friendly interfaces and automatic scoring methods. Indeed, paper and pencil tasks are still frequently used, and some of them can be very long-lasting. Moreover, the storage of documents related to the assessments is usually in paper-form, and a digitalization in this direction could result in a reduced environmental impact.

In this paper, we digitalised the paper and pencil test included in *Assessment of Pragmatic Abilities and Cognitive Substrates (APACS)* [3] developing a tablet application. The APACS Interview has a duration of 5 minutes in its original paper-pencil version, and it is made up of a single paper sheet in which the clinician registers on a checklist the frequency of occurrences of discourse anomalies with a score ranging from 0 to 2. The user-friendly digital interface of this application is expected to make the scoring procedure more flexible, providing an automatic score calculation that would be useful in clinical settings.

The paper is organized as follows: Section 2 presented the realted work. The complete system is described in Section 3 and Section 4 discussed some problems related to technological limitations. We conclude in Section 5.

2 RELATED WORK

Previous studies have shown successful results arising from interdisciplinary work between IT and both medical and cognitive science (e.g., [14], [15], [19], [21] and [20]). For example, Saxton [19] realized a Computer Assessment of Mild Cognitive Impairment (CAMCI), which is user-friendly computer task providing automatic scores. The authors compared CAMCI with the paper-and-pencil Mini-Mental State Examination (MMSE [10]), and checked the sensitivity of CAMCI in the identification of Mild Cognitive Impairment (MCI) in 524 nondemented individuals aged 60 years old. Their comparison between the two tests showed that the MMSE was almost insensitive to MCI, whereas the CAMCI provided high rates of sensitivity (86%) and specificity (94%). However, their study did not provide a reliable classification of MCI, which is important from the clinical point of view.

In another study, Singleton and colleagues [20] realised a Computer-based cognitive assessment of the development of reading, in children of 5 years old. Their computerized measure showed a highly satisfactory prediction of poor reading skills, with very low or zero rates for false positives and false negatives. However, it is important to consider that a word recognition test given at age 6 was not found to predict reading at age 8 with the same level of accuracy, resulting in an unsatisfactory false positive rate of 21%.

In the assessment of the language domain, the importance of the digitalization of cognitive tasks is emerging especially in discourse analysis, which requires a multidisciplinary work across Psychology, Linguistics and Information Technology [22]. Discourse production in pathological populations requires the evaluation of an extensive number of variables like repetitions, incomplete sentences, over- or under- informativeness of discourse, loss of verbal initiative, wrong order of the discourse elements, abrupt topic shift, but also abuse of gesture, altered intonation, fixed-facial expression, loss of eye-contact, etc [3] [16].

In this work, the aim is to convert the format of the paper-pencil interview included in Assessment of Pragmatic Abilities and Cognitive Substrates (APACS [3]), in a digital version. Although research has already been carried out the transformation of paper-pencil tasks into digital tasks [5], it is important to consider possible score discrepancies when comparing paper-pencil and the computer-based version [7].

To our knowledge, this is the first study that attempts to make the digital version of highly professional tool for the evaluation of several dimensions of pragmatic and paralinguistic features of discourse production.

3 DESCRIPTION OF THE APPLICATION

As already discussed in Section 1, APACS is a complex assessment tool composed by six tasks, two assessing production (i.e., Interview and Description) and four assessing comprehension (i.e., Narratives, Figurative Language 1, Humor, and Figurative Language 2). The "Interview task" consists in a spoken interview with the patient, with the goal to analyse his/her answers to simple questions. The interview is the only part of the APACS test which doesn't require the patient any particular action, the whole work is entrusted to the examiner. In particular, during the interview the examiner has to lead the talk, listen to the patient, recognize, count and remember patient's spoken anomalies of discourse, assign a mark to them, pay attention to interview's duration, and assist the patient when he/she has some difficulty or

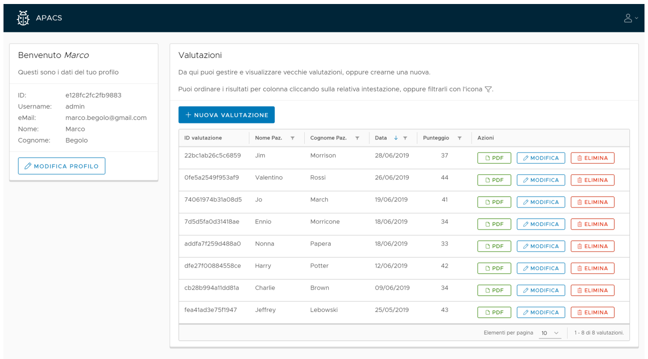

Figure 1: Application's dashboard, with personal information panel and list of exams panel

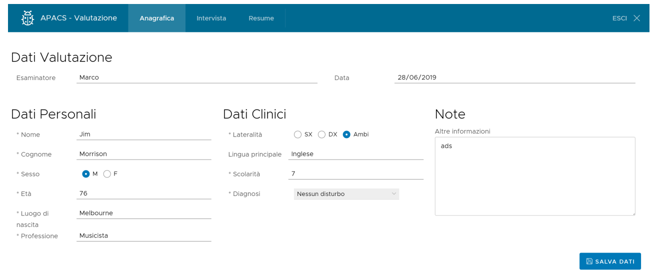

Figure 2: Patient's information form

is not inclined to speak. The total amount of points is 44, which gives an idea of the amount of information the examiner has to consider. All these tasks need to be performed simultaneously in a very short interval of time, only 5 minutes, resulting in an overwhelming job which may lead to inaccurate results or non-objective evaluations.

Often, eHealth mobile applications cover proactive contexts, in which the patient must actively perform some kind of task, or his/her actions are passively measured, recorded or elaborated, therefore the "user" overlaps with "the patient". Our application has the main goal to reduce examiner's stress in order to obtain more accurate evaluations, therefore in this case "the user" coincides with "the examiner" who is generally a doctor or a psychologist.

The application is composed by a server, which records data about patients, the audio registration of the interviews and the evaluation of the patient's performances, and by a tablet application that can be used by psychologists to record and review data.

The application initially provides a login page and a registration form, to grant access to doctors. Once logged in, each doctor/psychologist can access to his/her personal information, to all the previous administrations of the APACS Interview task or can record a new administration.

Figure 1 shows the dashboard that allows to perform all these three tasks. While the personal information section is just a standard form, to view and perform interviews the psychologist must create a new adminitration (i.e., "Nuova Valutazione" in Italian). During the interview the app provides an interface which aims to help the examiner to keep focused on features detections.

To avoid errors, the colours used in the interface change when the doctor create a new administration, in particular the bar on the top which is initially black (see Figure 1) becomes blue (see Figures 2 and 3). The examiner has to stop the recording, and exit from the interview panel, to go back to the dashboard. The scores can be corrected after the interview, but the idea is to avoid to accidental, not wanted, alterations of the scores.

The interface of the application is inspired to the original paper-and-pencil Interview in APACS test's pattern. The first panel is just a digitalization of a standard form for acquiring personal and clinical information of the patient, and the same holds for third panel, the resume, which is a simple resume of the data and scores acquired during administration. The interface of the interview is the most interesting, because it required a completely redesign to effectively help the examiner and reduce his/her stress (and, consequently, his/her error rate). We will discuss design choices in the following subsection.

The interview panel, depicted in Figure 3, consists in a big panel presenting all the possible anomalies of discourse as button elements, divided in 5 groups, with a point and click interaction. The examiner has to click on the correct button each time he/she discovers an anomaly of discourse (or just *anomaly* from now on). Each click counts the occurrence of the particular anomaly and eventually modifies the score for that specific clicked element. In this way, the examiner does not need to remember several numbers of occurrences of different anomalies of discourse to give a score at the end, so he/she is focused only in observing the patient to discover anomalies of discourse.

We consider two kind of anomalies of discourse, *countable* and *uncountable* and three possible scores, 2 (anomaly never occurred), 1 (anomaly occurred few times) or 0 (anomaly occurred many times, often or always). Initially, all the anomalies of discourse are scored to 2, i.e., never happened. Then, countable anomalies of discourse (like "Agrammatism") are scored to 1 the first time the their button is clicked. The second threshold is different for each specific anomaly (and can be configured), and specifies the number of occurrences needed to pass from the score 1 to 0.

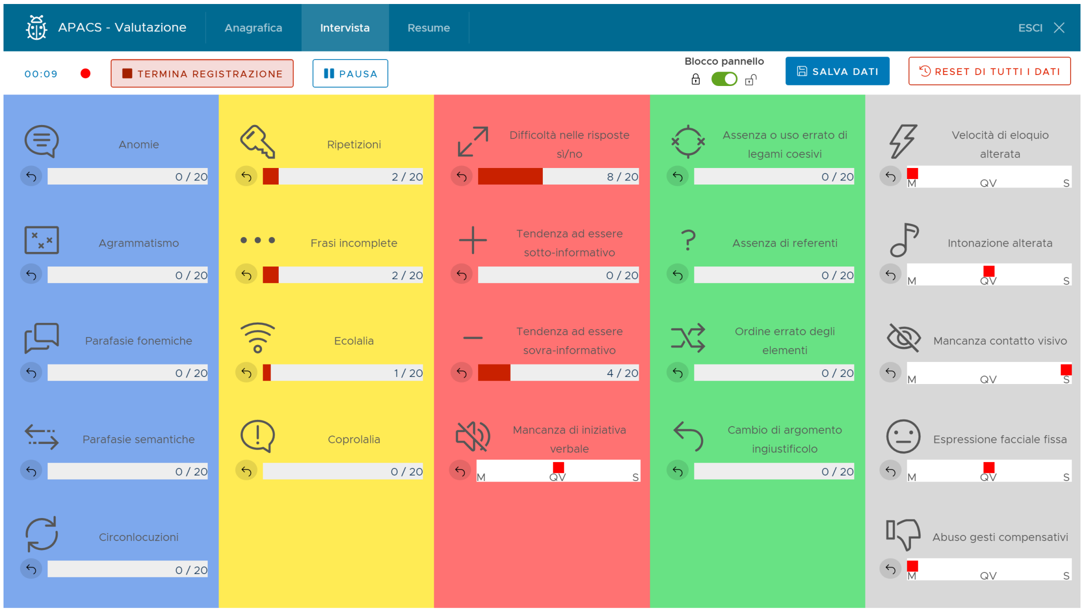

Figure 3: Screenshot of the interview panel

There are two reasons behind the choice of configurable thresholds: first of all, the examiner has no more to count in mind the number of occurrences, but he/she just need to react each time he/she recognizes an anomaly by clicking on the correct button, then the software is entitled to give the correct score. This is the main stress-reducing feature of the application, because it reduces the demand for the examiner of simultaneously calculate the frequencies and occurrences related to multiple discourse features while the patient speaks. This is why we designed a configurable threshold for each anomaly of discourse.

Both countable and uncountable elements are provided with an "undo" button, which allows to correct errors due to accidental click on the wrong button.

The interview panel results active (and its scores may be edited) only when a recording is performing. In this way, it has been possible to bind each exam to a recorded audio of the interview, in order to collect consistent data for future studies, possibly oriented to machine learning to achieve a fully automated classification of a patient's performance.

User Experience

The main idea behind the interview interface design comes from a videogame, *World of Warcraft*. In this game, there is a particular role that a player can plays, the role of the *healer*, in which the player has to focus on his/her group members, to take care of them, instead of on enemies. While the original UI of the game is good enough for single target situations, it is extremely inefficient for healers focusing on big teams of players (usually composed by 25 people or more). The task of the healer was to keep track of players' health and, when necessary, "shot an appropriate heal" (chose among many) on them.

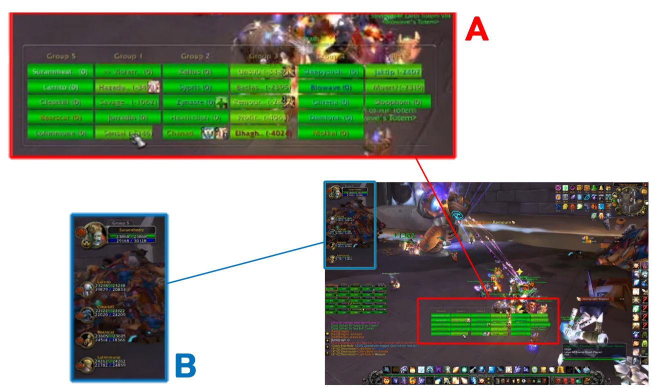

Figure 4: HealBot's UI in action: (A) HealBot's UI - (B) Original UI

This means that a healer has to simultaneously watch at 25 health-bars, give them a priority, then choose an appropriate spell and finally shot it on the correct player. All that in fractions of a second. This experience is usually really stressing and difficult. An interface add-on for this game, called "HealBot" is very useful since it provides a "point and click" featured team panel (see Figure 4), which is able to represent all team members (and their HP levels) and allows to shot a specific spell on a member using a combination of keys and a click performed on a player's tag. Switching to this add-on allows the player to reach top performances.

The task of the healers in this game is, in a certain sense, very similar to the one we want to target with our application. For this reason, we tested the point-and-click efficiency in a quick-reflexes context, and we adjust the design of our application according to the experience made with this game, exploiting how gamification could help to achieve better results in the administration of the APACS Interview task.

According to literature, since 2010 gamification has been widely adopted over the digital media industry and eHealt production [18], and also has been proved that it enhances user's visual attention and processing speed [13], so we decided that such kind of interaction would be optimal for this application. Nevertheless, gamification's benefits may be mitigated for new users who are new to the interface and don't know or remember button's role and position over the panel [11]. That's why we also minded the Fitts Law to design the application's interface.

It's well documented that Fitts Law regulates the scanning time T [9]:

$$T = a + b \log_2(\frac{D}{W} + 1)$$

where T is the interval of time needed to move from an object to another, W is button/icon's sizes (both height and width), D is the distance between buttons, and a and b are two empiric constant. With the aim to minimize the time needed for the movement, we decided to maximize buttons' and icons' sizes [W] (both height and width), and to minimize the distance between buttons by reducing margin among buttons, without going under the minimal distance of 2mm.

We distributed buttons equally on a grid of 5 columns (one for each anomaly's category) with 4 or 5 elements each, for a total of 22 buttons, each representing an anomaly in discourse production.

Furthermore, to help examiner's to recognize and memorize button's position on the grid, we provided an icon associated to each type of anomalies of discourse (in addition to anomaly's name), and we also chose 5 different colours to tag each column (category) differently [12]. We explicitly choose to not show category names over each column because they are not relevant for the assessment.

Technical Implementation

As already said, the system is composed by a server and a tablet application. Data collected during the interviews are stored in the server. We used a *REST* architecture developed in *node.js*[1] over a *MongoDB*[2] database.

[1]https://nodejs.org/en/

[2]https://www.mongodb.com/

Software	+	-	Total	Diff	Levenshtein Distance	% err
Google	27	62	89	35	214	30,18
Microsoft Azure	30	101	131	71	404	56,98
Trint	38	55	3	17	191	26,94

Table 1: Data about transcription experiments.

We chose *Deployd*[3] as a headless CMS server application to manage the Model, and we developed the client in *Angular 6*[4]. During the test phase the client was accessible as a web-application on a Nginx webserver, but the main advantage offered by Angular over a REST distributed application lays on the possibility to easily develop cross-platform clients starting from the web-application, thanks to frameworks like *Cordova*[5].

The database has been developed to gather data in order to create a consistent dataset suitable for supervised learning; for this reason, each administration consists in a set of scores associated to an audio recording of the interview performed by the examiner.

4 TECHNOLOGICAL LIMITATIONS

The first idea for this work was to create an application which was able to automatize the scoring activity for a group of selected anomalies. This is the reason why we record the interviews: we initially aim to analyse the audio recording to obtain a text transcription and then to analyse the text transcription to automatically find the anomalies. It was clear that not all the score can be calculated automatically, but we were confident in the success of task for some selected anomalies because, given a text, it is quite easy to search for repeated words, or for a set of bad words. Moreover, we planned to use natural language process techniques for more complex anomalies.

The problem arises when we collected the audio recording. We collected 20 recordings of interviews to old patients affected by Parkinson's disease and we used 3 services for transcription: Google, Microsoft Azure and Trint[6]. Unfortunately, none of the service was able to transcript the conversation with an acceptable percentage of errors. As an example, data about our experiments with one of the interview are reported in Table 1, where the columns "+" and "-" reported the number of words that are present in the transcriptions and not in the audio recordings and viceversa and we use the Levenshtein distance to calculate the difference

[3]https://deployd.com/

[4]https://angular.io/

[5]https://cordova.apache.org/

[6]http://trint.com

from the string obtained from one of the selected server and the correct transcription.

Other experiments highlighted that the number of errors decreases if there are only one person speaking, and not two (i. e. the psychologist and the patient) and if the patient is near the microphone. Moreover, we noted that our patients does not only have often difficulties in speech production, but also they usually speak in dialect, which is widespread among Italian elders, but it is a real problem for speech recognition softwares which can be used only with national language.

For this reason, it was not possible to obtain a text transcription of the interviews, and we cannot realize our initial project.

5 CONCLUSIONS

In this paper, we describe a system for assisting the psychologists, during the administration of the Interview task, included APACS. The tablet application provides a user-friendly interface that helps to focus with more accuracy on the detection of possible anomalies. To our knowledge, this is the first study that attempts to make the digital version of tool that evaluates several dimensions of pragmatic and paralinguistic features of discourse.

From the point of view of the examiner, who is the effective user of our system, this is a very useful and innovative way to assess patients' cognitive performance. In the specific case of APACS Interview, the possibility to interact with the application in a user-friendly way allows the clinician to allocate more attention to what the patient is saying, instead of keeping "on-line" the frequencies of occurrences of discourse anomalies, when they occur.

Moreover, the automatization of global score calculation has the advantage of reducing possible errors that might naturally derive from manual procedures.

Another important consideration refers to the fact that performing the task in a digital form, and then obtaining a PDF final summary (with information of the patient plus his/her cognitive outcome), may subsequently lead to limited paper consumption, which in turn may results in a reduced environmental impact.

REFERENCES

[1] M. Amanzio, G. Geminiani, D. Leotta, and S. Cappa. Metaphor comprehension in alzheimer's disease: Novelty matters. *Brain and Language*, 107(1):1–10, 2008.

[2] M. Amanzio, G. Geminiani, D. Leotta, and S. Cappa. Communicative-pragmatic disorders in traumatic brain injury: The role of theory of mind and executive functions. *Brain and Language*, 168:73–83, 2017.

[3] G. Arcara and V. Bambini. A test for the assessment of pragmatic abilities and cognitive substrates (APACS): Normative data and psychometric properties. *Frontiers in psychology*, 70(7), 2016.

[4] V. Bambini, G. Arcara, M. Bechi, M. Buonocore, R. Cavallaro, and M. Bosia. The communicative impairment as a core feature of schizophrenia: Frequency of pragmatic deficit, cognitive substrates, and relation with quality of life. *Comprehensive Psychiatry*, 71:106–120, 2016.

[5] S. Bertoni, A. Facoetti, S. Franceschini, C. E. Palazzi, and D. Ronzani. A web application for reading and attentional assessments. In *Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good (GOODTECHS)*, pages 142–147, Bologna, Italy, November 2018.

[6] A. Carotenuto, G. Arcara, G. Orefice, I. Cerillo, V. Giannino, M. Rasulo, R. Iodice, and V. Bambini. Communication in multiple sclerosis: Pragmatic deficit and its relation with cognition and social cognition. *Archives of Clinical Neuropsychology*, 33(22):194–205, 2018.

[7] Collerton, J. and Collerton, D. and Arai, Y. and Barrass, K. and Eccles, M. and Jagger, C. and Newcastle 85+ Study Core Team. A comparison of computerized and pencil-and-paper tasks in assessing cognitive function in community-dwelling older people in the newcastle 85+ pilot study. *Journal of the American Geriatrics Society*, 55(10):1630–1635, 2007.

[8] A. L. Fernandez. Modern neuropsychological tests for a diversity of cultural contexts. *The Clinical Neuropsychologist*, 33(3):438–445, 2019.

[9] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47:381–391, 1954.

[10] M. F. Folstein, S. E. Folstein, and P. R. McHugh. "mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, 12(3):189–198, 1975.

[11] Morten Hertzum and Kasper Hornbæk. The effect of target precuing on pointing with mouse and touchpad. *International Journal of Human-Computer Interaction*, 29:338–350, 2013.

[12] Craig Hockenberry. *Making sense of color management.* A book apart, 2016.

[13] Lumsden J, Edwards EA, Lawrence NS, Coyle D, and Munafò MR. Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. *JMIR Serious Games 2016;4(2):e11*, 2016.

[14] D. D. Kumar, S. L. Helgeson, , and A. L. White. Computer technology-cognitive psychology interface and science performance assessment. *Educational Technology Research and Development*, 42(4):6–16, 1994.

[15] A. W. Kushniruk, V. L. Patel, , and J. J. Cimino. Usability testing in medical informatics: cognitive approaches to evaluation of information systems and user interfaces. In *Proceedings of the AMIA annual fall symposium*, page 218. American Medical Informatics Association, 1997.

[16] A. Marini, S. Andreetta, S. Del Tin, and S. Carlomagno. A multi-level approach to the analysis of narrative language in aphasia. *Aphasiology*, 25(11):1372–1392, 2011.

[17] S. Montemurro, S. Mondini, M. Signorini, V. Bambini, and G. Arcara. Disruption of pragmatic abilities in parkinson's disease: the role of cognitive reserve. In *Proceedings of XPRAG.it 2018 - Second Experimental Pragmatics in Italy Conference*, 2018.

[18] Lamyae Sardi, Ali Idri, and José Luis Fernández-Alemán. A systematic review of gamification in e-health. *Journal of Biomedical Informatics*, 71:31 – 48, 2017.

[19] J. Saxton, L. Morrow, A. Eschman, G. Archer, J. Luther, and A. Zuccolotto. Computer assessment of mild cognitive impairment. *Postgraduate medicine*, 121(2):177–185, 2009.

[20] C. Singleton, K. Thomas, and J. Horne. Computer-based cognitive assessment and the development of reading. *Journal of Research in Reading*, 23(2):158–180, 2000.

[21] B. Smith and P. Caputi. Cognitive interference model of computer anxiety: Implications for computer-based assessment. *Computers in Human Behavior*, 23(3):1481–1498, 2007.

[22] T. A. Van Dijk. *Discourse studies: A multidisciplinary introduction.* Sage Publications, 2011.