

Adapting MAIN to eliciting stories from adults and in a remote context: What do we have to consider, and what do we know?

Katrin Bente Karl^{*1}

University of Bern, Switzerland

The present study suggests guidelines for the successful elicitation of adults' narratives using an online remote design. In doing so, I have adapted the Multilingual Assessment Instrument for Narratives (LITMUS MAIN) to an adult population and specify possible applications in a remote context. Hereby, I elaborate on various features that impact the elicited data and the testing context. I also report results from a pilot study with 10 adults telling MAIN stories using three different testing methods (two moderated methods using PowerPoint or an external link and one unmoderated) to argue that different methods of remote narrative elicitation do not influence the macrostructure of the narratives. However, by extending the analysis to the context of the testing and including the experimenters' and participants' assessments of the testing situation, we can see differences that lead me to recommend the so-called Link method (a moderated remote testing method with a certain degree of autonomy) for remote testing with adults.

1 Introduction

Recent developments in data elicitation methods have moved toward the use of digital online (remote)² elicitation. Consequently, the use of several questionnaires and experimental data elicitation platforms is increasing (e.g., Psychopy, Pavlovia, jsPsych, Labvanced, and Gorilla, to name a few). Apart from this development, oral elicitation in-person became impossible during the years of the Covid-19 pandemic. This situation has led to proliferated methodological studies on data elicitation and fundamental empirical research based on online elicited data (e.g., the research topic Remote Online Language Assessment: Eliciting Discourse from Children and Adults in *Frontiers in Psychology*, cf. frontiersin.org). Among several digital tools for elicitation of lexicon and grammar, the existing instrument on narrative elicitation, Multilingual Assessment Instrument for Narratives (LITMUS MAIN; hereafter

* Corresponding author: katrin.karl@unibe.ch

¹ Fruitful discussions with Natalia Gagarina accompanied the design and completion of the study.

² I will use the terms online and remote synonymously throughout the paper.

MAIN, Gagarina et al., 2012; Gagarina, Klop et al., 2019), also went online (see the different versions for online MAIN testing with children, which are available for download via the MAIN-homepage: the Hong Kong Polytechnic University – ZAS Version, Slovak version and ZAS version, <https://main.leibniz-zas.de>). Most studies using MAIN have elicited data from children (to name a few recent studies: Fiani et al., 2022; Roch & Hržica; 2020; Rodina, 2020; Sheng et al., 2020, and the contributions in Gagarina & Bohnacker, 2022a, b). However, there is growing interest in transferring the established data collection method to adults (cf. Gagarina et al., 2019b; Krasnoshchekova & Kashleva, 2019) and in a digital context (cf. Jažić et al., 2023; Mattiuada et al., 2022).

In this paper, I elaborate on the cornerstone criteria for successful online narrative elicitation from adults. I suggest an updated method for working with adults using MAIN. I start with outlining a number of fundamental criteria to consider during the test situation (Section 1) and move on to the adaptation to an adult context (Section 2). Next, I outline the adapted instructions for eliciting MAIN with adults (Section 3) and illustrate different testing methods (Section 4). Finally, I present the design, implementation, analysis, and results of a pilot study (Section 5), and conclude with a final discussion (Section 6).

2 General remarks: Test situation

The multidimensional theoretical model of narrative text organization laid the foundation for the MAIN tool. It is based on the consideration that a narrative contains microstructural (language-specific linguistic structures at all linguistic levels) and macrostructural elements (hierarchical organization of the story, episodic structure, story grammar) that need to be analyzed separately, but can be collected and examined within a framework (cf. Gagarina et al., 2012, p. 8f.). Concerning the macrostructure, MAIN allows for the measurement of three elements: story structure (following the story grammar model, cf. Stein & Glenn, 1979), which claims that a narrative consists of components, including the setting of the story and episodes), structural complexity (following Westby, 2005 and taking into account that an episode itself consists of components, such as Goal, Attempt, and Outcome and can be measured by its level of structural complexity and divided into sequences, incomplete and complete episodes) and Internal state terms (showing to which degree the narrator emphasizes with others' emotions and state of minds, as it is elaborated in the theory of mind (ToM)) (cf. Gagarina et al., 2012, p. 10ff.). The MAIN protocol (Gagarina et al., 2012) allows for the analysis of the verbalized items of a narrative collected with the help of MAIN. Each item scores one point. For the entire narrative, by verbalizing all components of all episodes, all internal states, and the setting of the story, a maximum of 17 points can be achieved in total.

The foundation of this model, the causal-temporal episodic components, is universal for narrative organization. Its pictorial representation was originally developed for children between the ages of 3 and 10 (cf. Gagarina, Klop et al., 2019, p. 1) but later was successfully used with teenagers and adults (Antonijević et al., 2022; Gagarina et al., 2019b; Jažić et al., 2023; Krasnoshchekova & Kashleva, 2019; Mattiuada et al., 2022; Vogels & Lindgren, 2022). However, it also requires a background view, which I want to provide here.

First, it is necessary to take a closer look at the test situation, which includes the participant, experimenter, the picture-based story, and their respective interactions. From the moment of narration, we also deal with the roles of narrator and addressee. In the test situation described in the MAIN protocol (for testing children, cf. Gagarina, Klop et al., 2019, p. 3), the experimenter is the same as the addressee of the story and the participant the narrator, which leads to a child narrator telling an adult a story (Figure 1).

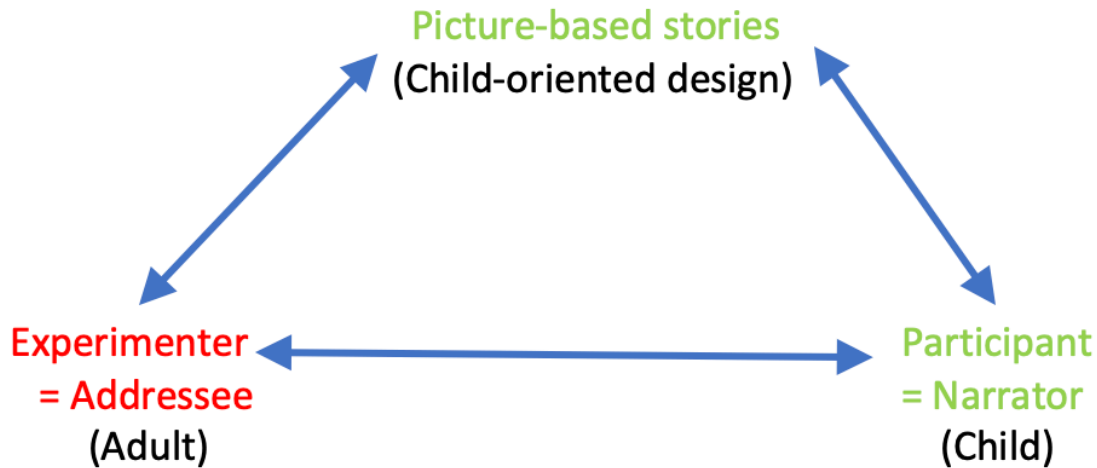


Figure 1: Graphical representation of the situation of testing with children.

In this constellation, the conception and design of the story and the participant have the same (child) context. In contrast, the story's addressee (experimenter) is an adult. Depending on the child's experience of telling picture-based stories to an addressee, be it adults or other children, the test situation can be more or less familiar to the child.

While transferring the elicitation method to an adult context, the participant changes from a child to an adult, as shown in Figure 2.

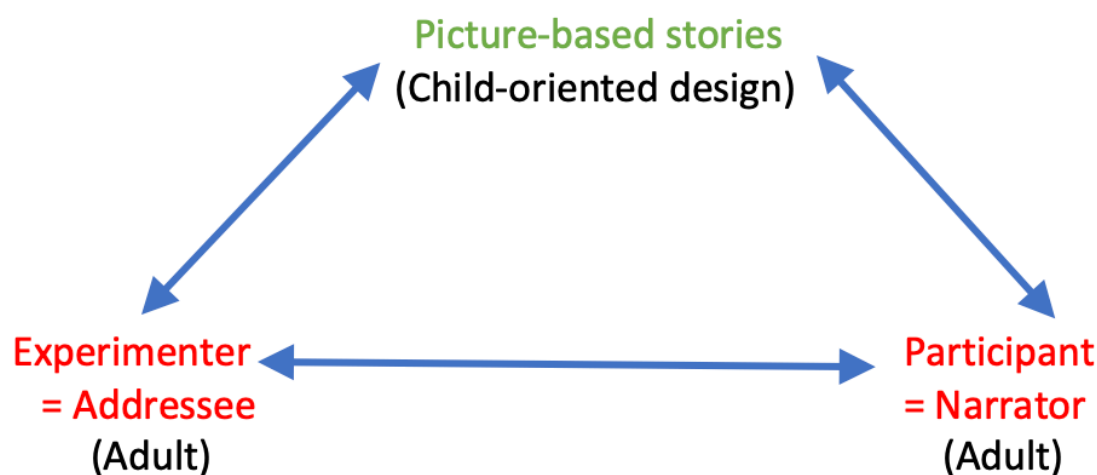


Figure 2: Graphical representation of adult testing.

Here, the experimenter and the participant are both adults, whereas the material used remains in the child's context. Similarly to testing children, there is thus a contextual difference (adult

vs. child) in one component, while the other two match. What consequences this has requires a more thorough discussion.

3 Adult participant (consequences for testing)

The term *adult* is used with different meanings. In everyday language, it can refer to a person who has reached a) a biological state (i.e., physical maturity, being grown up), b) a legally defined age with which the attainment of maturity goes hand in hand, or c) a cognitive maturity (Oxford English Dictionary, www.oed.com). It does not refer to a division into stages, nor does it have an endpoint. It is an elastic and thus also imprecise term; a person is considered an adult at 20 and 90. Ultimately, the term adult can distinguish a person from a child or a teenager. In this general sense, I refer to adults as mature persons who are no longer in the process of growth and – probably the most relevant point for our interests – have completed their first language acquisition. This vague definition requires further explication. In current research, there is a broad discussion on the classification of adulthood (e.g., adolescence, early, middle, and late adulthood), as well as on changes during adulthood (e.g., concerning personality, see Srivastava et al., 2003, or language across the lifespan, see Beaman & Buchstaller, 2021; de Bot & Schrauf, 2009; Gerstenberg & Voeste, 2015). The changes people undergo are not precisely predictable; they depend on individual factors, such as lifestyle, socioeconomic status, and biological components (e.g., cognitive or general health peculiarities), which we must consider. The necessary extent of background information and its relevance in concrete studies depends on the research question and the study design and thus cannot be conclusively clarified in this article. Since my considerations are concerned with the general difference between children and adults (and thus with the very general distinction described above), I assume an adult who has completed his first education and first language acquisition and who shows no cognitive limitations.³ No general age limit is to be set here. Such an average adult differs from a child in many respects. For the context of these remarks, such appear relevant as, e.g., the life experience and perhaps higher experience in test situations; in our time, likewise a basic familiarity with technical devices, software, and media;⁴ acquired habits and routines in storytelling and dealing with different communicative situations and reacting to materials (as the stimuli for narrating a story) and so on. In the following, I will elaborate on some of these points.

3.1 Children vs. adults in testing situations and interaction with the experimenter

In a testing situation we can see a difference between testing children and adults on two levels – how a childish vs. an adult participant deals with and reacts in the testing situation in general and how they interact with the experimenter. Concerning the first issue, the research from Bauer et al. (2017) shows that performance in a testing situation differs consistently between children

³ These conditions can thus not be transferred to the testing of older people and people with different cognitive impairments. For testing those people, separate considerations are inevitable. For testing with adults with down syndrome cf. Mattiauda et al. (2022).

⁴ That also may apply to children and, to a limited extent, to older people.

and adults,⁵ stating that “adult-like performance was observed around mid-adolescence, by ages 13-16 years“ (ibid., p. 1), and that adults outperformed children in their experiments. Punch (2002) explains in more detail the question of to what extent research with children differs from adults and gives advice to experimenters testing children. Although her considerations take the opposite view (deducing from testing with adults to testing with children), they provide insights into the main differences as she states that a crucial point is the perception of children vs. adults. And there, we can see the link to the second issue, the interaction between the participant and the experimenter.

If we look at the above interaction scheme (Figure 2), we see a relevant difference regarding the interaction between the participant and the experimenter. In our case, two adults interact with each other and not an adult with a child. Consequently, we have a difference in address, which becomes apparent both linguistically (according to the respective linguistic conventions, e.g., adaptation of the distance and/or polite address, which, of course, is highly dependent on the concrete languages in question) and pragmatically. An adult may be interested in the context of the test situation and may have previous experience with testing. The experimenter should address these issues and explain the background of the test and research interest to a different extent than is the case with especially younger children.

3.2 Participant: Skills, habits, and digital literacy

As addressed above, children and adults may differ in skills, habits, and, potentially, digital literacy. For this article, we have to take a closer look at the narrative skills, habits in telling a story, and the assumed level of proficiency, to handle different testing software.

Narrative skills can be understood as the ability to derive a meaningful context from a sequence of events and to verbalize this context. To be able to narrate, we have to have some preconditions. Pfeffer (2015) points to cognitive skills, such as working memory and processing capacity, emotional or pragmatic skills (c.f. the already mentioned theory of mind and the possibility of emphasizing with the listener) and, finally, linguistic skills to be able to verbalize the story. These skills show up in narratives of different kinds, amongst them picture-based stories. It is considered a skill that is acquired late in language acquisition. Research using the MAIN stories show that macrostructure develops with children’s age (see Gagarina et al., 2019b, p. 195 and the cited literature), but less is known yet about developmental changes in adults (is there an endpoint of acquisition, is the skill – once acquired – stable or which factors could possibly affect their development?). These questions cannot be explored in depth or answered in this article but must be discussed in follow-up studies (as planned in Karl, in prep.). The relevant point is that narrative skills in adults are no longer in the acquisition process, as is the case with children.

Regarding individual habits in telling a story (e.g., when/in which situation, to whom, and how regularly the person tells stories), there are differences between children and adults, but they can appear as well as within adults considering their stage of life, routines, and

⁵ I want to point out that these studies mainly refer to generalized adults and do not refer to possible age differences. The extent to which test-taking behavior changes throughout adult life is the subject of other studies (cf. Whitbourne, 1976, or Andreoletti et al., 2006 for measuring anxiety).

interests. It is plausible that these factors influence a person's narrative. As they are highly individual and depend on life experience and personal habits, they cannot be generalized. Still, information on them should be gathered when it comes to concrete testing.

Finally, I want to come to digital literacy, which can be considered an “umbrella term” (Nascimbeni & Vosloo, 2019, p. 10) and shall be restricted for the purpose of this article to “the ability to use digital devices or software” (ibid.). Especially for younger adults, the use of computers, the Internet, various software, and, at the latest, since the pandemic, platforms such as Zoom are part of everyday life and hardly pose a challenge. Therefore, we can assume that they have a relatively high digital literacy, which allows for testing them in different remote settings (see Section 5.1 below). However, the older the participants become, the more carefully one should look at the extent to which they are familiar with technical devices (especially at an older age) and adjust the test design in case of doubt.

3.3 Story and addressee

The final important component to consider is the story to be told (the material). In our context, this is a fixed factor, since in order to compare narratives from children and adults, the same picture-based stimuli need to be used. However, it still gives rise to the consideration of two points: the child-oriented material and the addressee of the story.

Concerning the material, I advise contextualizing the materials used in advance in an adult-specific way by adding a short description of why these materials are used: Such a description could be phrased in the following way: “We use test materials established for collecting stories from children. Therefore, they are adapted to children according to the type and structure of the stories. Now we are interested in how the same stories are told by adults of different age groups.” Such an explanation can serve as a justification for the participant.⁶

The question of the addressee concerns a core point of the test design and leads to a more detailed discussion in the following section.

4 Instructions for testing MAIN with adults

Against the background of the previous remarks and considering the extensive research on eliciting data with the help of MAIN with children (for an overview, see Lindgren et al., this volume), two points are particularly relevant for developing instructions for testing the MAIN stories with adults: the aim of gaining as much comparability as possible with the established procedure and the need to adapt some points to an adult context at the same time. These two demands are not contradictory, which is why a transfer from a child to an adult context is generally unproblematic. However, there is a need to discuss the points already mentioned above: interaction with the participant (including addressing and considering the adult's abilities), choice of testing method, and finally, the question of the addressee of the story. These three points influence the test situation and, thus, the data quality differently. In a presentation by Karl (2021), there was a discussion regarding the participant. There is a need to consider the respective linguistic politeness norms of the tested language and the individual relationship

⁶ Regarding testing with older children or adolescents, one could also consider adding this explanation.

between the experimenter and the participant (if I know the person, it would be unnatural to address him or her differently than in a common communication situation, for example in the distance form during the test, just because it is fixed in the manual). Likewise, a more detailed presentation of the test background, including a reference to the origin of the visual materials, could be included in the instructions without significant trouble. The question of the testing method seems much more relevant as it could potentially influence data quality.

The MAIN protocol was initially devoted for in-person testing with children but has already been adapted for a remote context. Different materials for online-testing are available via the homepage (see <https://main.leibniz-zas.de/en/main-materials/main-materials/> and Section 1 above). Studies comparing elicited narratives with children in both contexts show comparable results regarding macrostructure in production and comprehension (cf. Pratt et al., 2022 and the special issue of *Frontiers in Psychology* on digital discourse elicitation).

As mentioned above, MAIN has been used with adults employing an in-person testing procedure in a study by Gagarina, Bohnacker et al. (2019b). In this study, 69 monolingual German-, Russian-, and Swedish-speaking adults told one of the MAIN stories (either Baby Birds or Baby Goats). Among others, Antonia Hannes accomplished a transfer into a remote context of testing with adults. In her Bachelor thesis (2021), she tested 20 German-speaking adults using a Zoom mode of testing. The MAIN stories were presented via a PowerPoint presentation operated by the experimenter (using screen sharing) with the help of the corresponding material for online child testing. The data collected in this way showed no significant deviations in macrostructure from the data collected in the in-person testing with adults from Gagarina, Bohnacker et al. (2019). Thus, there is essential comparability. In addition, there are studies collecting data from adults and using them as a control group (e.g., Hržica & Kuvač Kraljević, 2022; Vogels & Lindgren, 2022) and two important studies focusing on adult's narratives elicited with the help of MAIN in a remote context (Antonijević et al., 2022; Jažić et al., 2023). Antonijević et al. (2022) aimed to establish measures of macrostructure in narratives of Irish-English bilingual adults to create a baseline for comparison with narratives of respectively bilingual children. Therefore, they tested 30 Irish-English speaking adults (aged 22–59 years) via Zoom with the help of an unpublished “custom-made PowerPoint presentation embedding the 6 pictures for each story [...] [to] share the pictures with participants.” (ibid. p. 5). Regarding the MAIN protocol (Gagarina, Klop et al., 2019), in the beginning, they presented three envelopes to “create an illusion that the researcher did not know the story that the participant was about to tell.” (Antonijević et al., 2022, p. 5). Concerning the online administration of MAIN, they hint at some advantages, such as the experimenter's evaluation of straightforwardness, an improvement in time management and flexibility but also mention a possible malfunction of the internet connection and, as a limitation, the small or differing size of the pictures (ibid. p. 9f.).⁷

Jažić et al. (2023) also uses MAIN stories in a remote context for gathering narratives from adults, in their case, from 20 monolingual and 20 heritage speakers of Bosnian, aged between 18 and 30 years. They aimed to research differences in these two populations regarding

⁷ Pratt et al. (2022) hinted at that too.

case marking. Hence, they focus on microstructural features and use the MAIN stories to elicit data with the help of a PowerPoint Presentation and using Zoom. They follow the here to be yet elaborated procedure (see below) and use the material developed for online testing with adults (*ibid.*, p. 14). So, this study is of great interest as it shows that the developed material can be used successfully. Furthermore, they hint at the same crucial points, which were also discussed during an online meeting on 26.03.2021 and will be described in more detail in the following.

The first point concerns the need for discussion on the specific design of the PowerPoint slides, especially concerning how to deal with the fact that the participant is supposed to tell the story without the experimenter knowing which story it will be (non-shared knowledge), as specified in the MAIN standardized procedure. The choice between the three different envelopes and the (supposedly) random selection in the in-person testing for children grant this. This experimental setup is also possible in an in-person testing with adults. However, by presenting the story via screen sharing with the help of PowerPoint, one can hardly assume that an averagely skilled adult participant believes that the experimenter cannot see the pictures just shown and accordingly tells the story in non-shared mode (see the notes on that in Jažić et al. 2023, p. 8).

The next question arises from the adequate addressee of the story. The visual material (developed for children) makes it seem logical to give adults the task of telling this story to an (imaginary) child. In this way, the discrepancy between testing an adult with the child-oriented material already raised could be solved well. However, if one decides to do this, a vital difference emerges compared to the instruction for children who tell this story to adults. Consequently, the comparability of the data becomes questionable. In addition, before changing the addressee we must know more about the effects on the macro- and microstructure of the stories told. Systematic studies on this topic have yet to be conducted, so it is wiser not to carry out such a change at this point. Instead, it seems sensible to first conduct contrastive studies with varying addressees (Karl, in prep.).

To sum up, remote instruction should initially include addressing the story to an adult, and the question of non-shared knowledge needs to be solved differently than in the case of child testing.

Considering these arguments and in exchange with other researchers from the MAIN-community (a.o., Natalia Gagarina & Anna Smirnova, exchange during an online-meeting on 26.3.2021), the slides developed for online testing children (the so-called ZAS version) were adapted by including the considerations mentioned above. To enable non-shared knowledge, the adults' ability of imaginativeness is used: they shall imagine that the experimenter does not see the pictures during the narration. Hence, we ensure parallelism and consider adults' competencies (their ability to imagine things). These reflections were incorporated into the PowerPoint slides and implemented as simply as possible (for an application example see Jažić et al., 2023). The materials are thus oriented as closely as possible to the in-person test materials and only differ from them to the extent necessary. The developed PowerPoint slides and instructions (for online and in-person testing) for testing with adults will soon be available on zenodo (the author can be contacted for more information).

Thus, these adapted materials are a first outcome. At the same time, a number of questions remain open: the question of enabling non-shared knowledge in remote testing and the effects of addressing the story to an adult or a child. In the following section, I will discuss the first question in more detail. For answering the second question, I refer to a comprehensive ongoing study which systematically investigates this variation and its effects (Karl, in prep.).

5 Elicitation methods

5.1 General remarks

We have to distinguish between two general testing settings to elicit data: The in-person testing and different forms of digital remote testing. Both can be further differentiated and are referred to with varying terms (e.g., remote/online testing, computer-based testing (CBT) vs. in-person testing, paper-based or paper-and-pencil tests). There are a growing number of studies comparing these two testing settings and discussing the differences in general and regarding the outcomes (to name a few more recent studies from different disciplines: Funda, 2021; Karay et al., 2015; McClelland & Cuevas, 2020; Smolinsky, 2020; or for a broader overview Leeson, 2006). For the purpose of this article, I want to give a very broad overview of some of the relevant differences. The following variables may differ between the methods of testing:

- Space (experimenter and participant are/are not present in the same room)
- Time (experimenter and participant are/are not present at the same time)
- Moderation (a real person does/does not moderate the test)

In a classic in-person test situation, the experimenter and the participant are in the same room simultaneously, and a real person (experimenter) moderates the test. In contrast, by definition we have a non-shared real space in all remote testing procedures. However, in a typical situation, the persons meet in a digital space where both are present at the same time. The test is still carried out in the presence and moderated by an experimenter. I will call these tests moderated remote testing (abbreviated as remote testing A).

Testing software makes it possible to conduct a test digitally at every possible moment without a real moderator. Such tests are called unmoderated or sometimes self-moderated (for a comparison between moderated and unmoderated tests c.f. a.o. Hertzum et al., 2015, for unmoderated testing He, 2021; Relawati & Primanda, 2022) (abbreviated here as remote testing B). Such a testing method differs from in-person testing for all variables. Table 1 summarizes the differences between the three methods, in-person testing, remote testing A and remote testing B.

Table 1: Methods of testing and its variables.

Methods of testing	Space	Time	Moderation
In-person testing	+	+	+
Remote testing A	-	+	+
Remote testing B	-	-	-

5.2 *Digital testing procedure in MAIN*

In order to adapt MAIN to remote testing procedures, we have to take into account the *degree of non-shared knowledge*, and the *degree of autonomy* of the participant. By the degree of autonomy, I understand the extent to which the participant goes through the test in a self-determined way and can, for example, influence the speed of presentation of the individual pictures. These two variables depend on the specific test instructions (i.e., not on the choice of method per se).

In canonical in-person testing (as described in the MAIN protocol), the choice of envelope allows for autonomy. However, the experimenter does the unfolding (and refolding) of the story or pictures. That reduces the degree of autonomy but guarantees the non-shared-knowledge effect maintenance: the experimenter can control that he/she does not see the pictures him/herself and can pretend to not know the story. In this testing procedure, the participant tells the story to the experimenter, a real adult addressee. The whole procedure is carefully considered, piloted, adapted, and finally recorded and adopted in this form for most tests with children and the case of adult testing already mentioned.

All remote testing procedures developed so far for testing with children use moderated testing in digital space (via conferencing platforms, such as Zoom) (remote testing A) with the help of a PowerPoint presentation. The experimenter starts the screen sharing and shows the MAIN story as part of a PowerPoint presentation. The experimenter navigates through the presentation; the participant has no access rights. In this variant, it is logical that the participant and the experimenter have the same view of the pictures, which raises the question of how the non-shared knowledge effect can be guaranteed. Therefore, different approaches exist for testing children. One version (Slovak) suggests to the child that the computer screens show different images – the experimenter’s screen remains black during the presentation of the pictures, which are only visible on the subject’s screen. Another version (Hong Kong Polytechnic University – ZAS Version, for both versions, see <https://main.leibniz-zas.de/en/main-materials/main-materials/>) introduces an additional imaginary protagonist and addressee (a child) in the presentation. Hence, there is a virtual (and additionally childish) addressee.

If we transfer this to the adult context, I again hint at the mentioned differences between children and adults. If we imagine an adult with a correspondingly high level of digital literacy, it would be difficult to trick him/her into thinking that the experimenter cannot see the screen during a test via Zoom and screen sharing (cf. also the comments of Jažić et al., 2023, p. 8). The second variant, introducing an imaginary childish addressee, also poses difficulties, and the deviation from an adult addressee to a child needs to be revised in terms of its effects. This gives rise to the following questions:

- 1) Is a parallel test design using a PowerPoint presentation (henceforth PPP) and screen sharing suitable for adults, and what modifications are necessary?
- 2) What other test designs are available in remote testing?
- 3) What impact do different test designs have on the collected narrative data?
- 4) Which test design is the most suitable in which context?

I explored these questions by conducting a pilot study in which I elicited narrations from adults using three MAIN stories in the telling mode (see Gagarina, Klop et al., 2019 for descriptions of the telling, retelling, and model story modes).

6 Pilot study

I conducted a pilot study with 10 German-speaking adults telling three MAIN stories (Baby Birds, Baby Goats, and Cat) in the telling mode using three different testing methods (two moderated methods – one using PowerPoint and the other an external link – and one unmoderated) to research their possible effects on different levels. These are the level of the narratives' macrostructure (Section 5.3.1, defined and analyzed following to the MAIN protocol, Gagarina, Klop et al., 2019) and the levels of assessment from the view of the participant (Section 5.3.2) and the experimenters (Section 5.3.3). The conclusion (Section 5.3.4) weighs the results and comes to the outcome to recommend the moderated method with the help of an external link for remote testing with adults.

6.1 Testing methods

Taking into account the premises for the transfer to the adult context (comparability, avoidance of unnecessary deviations, and retention of an adult addressee, see above), the following three methods for remote testing were developed:

1) Moderated remote testing via Power point presentation (PPP): For this procedure, the Power point slides revised for adults (see explanations above) were used via screen sharing in virtual meetings. The non-shared knowledge is guaranteed using the adult's ability to imagine (see above). We asked the participant, "Now, please tell the story. Look at the pictures, and try to tell the best story you can. While doing this, imagine that I cannot see the pictures. Please start your story." This procedure assumes that the participant can use the conference software, but no further digital literacies are necessary. The conference software records the narrations, and the experimenter moderates the test.

2) Moderated remote testing via software (hereafter *Link*): This method works via a survey software in which we embedded the images of the MAIN stories. A digital survey was created that is accessible via a uniform resource locator (URL), i.e., a link to a website. This link is shared with the participant via chat during a virtual meeting. The person gets access to the survey page and can navigate autonomously through the pictures of the story. Here, the introduction, with the three envelopes' selection and the pictures' presentation, is identical to the PPP version. The difference is that only the participant sees the story. This solves the problem of non-shared knowledge. At the same time, he/she can go through the pictures autonomously at his/her speed. With the appropriate screen setting (or two screens), the experimenter's view remains large enough to ensure good interaction. This implementation assumes that the participant can handle with different software and is familiar with using parallel windows on the screen. The conference software records the narrations, and the experimenter moderates the test.

3) Unmoderated testing: No meeting with an experimenter is necessary for this method. The testing is done autonomously via software, which is also accessible via a link. The participant receives the link, calls up the page, and goes through the survey with the help of videos, audio files, and written instructions. The participant navigates autonomously through the story and has the task of telling it to a virtual person. In this case, autonomy is comparable to the second variant. The question of non-shared knowledge does not arise because no addressee is present (nevertheless, the instruction was kept identical in wording in all cases). The significant difference is that no moderated testing takes place. Accordingly, there is no interaction with the experimenter and no opportunity to ask questions. The interview software records the narrations. The recording must be started by the participant himself/herself. To implement this test version, the highest digital literacy is necessary. The participant has to start the software on its own and has no possibility to ask for help.

For the design of the Link version and the unmoderated testing and thus for the transfer into digital space with the help of survey software, I obtained permission from MAIN authors (Daleen Klop, Sari Kunnari, Koula Tantele, Taina Välimaa, Ute Bohnacker, and Joel Walters). I chose the survey platform *soscisurvey.de*, as it complies with data security standards on the one hand and blocks the download of embedded images on the other hand – in this case, the MAIN stories – and thus offers copy protection.

A summarizing comparison of the mentioned testing methods is shown in Table 2.

Table 2: Testing methods and other characteristics.

Testing method	Space	Time	Moderation	Autonomy ⁸	Non-shared knowledge	Addressee
In-person testing	+	+	+	-	+	real
PPP	-	+	+	-	-(imagined)	real
Link	-	+	+	+	+	real
Unmoderated testing	-	-	-	+	not necessary	imagined

6.2 Participants and procedure

For the pilot study, I collected data from 10 adults aged 22 – 25 (mean age: 21.9) from the same socioeconomic background: They were all students (of different disciplines at a German university), with German as their first language. The testing begun with a written questionnaire with questions on biographical data and several items on linguistic habits, such as storytelling routines and contact with children. Afterward, the virtual meeting with the eliciting of the MAIN narratives took place. Each person told their three stories with the help of the three mentioned remote testing methods, i.e., each participant had to tell three different MAIN stories, all in the telling mode, each using another testing method. The stories used were Baby Birds, Baby Goats, and Cat in random order for the different testing methods, where either Baby Birds or Baby Goats was always the first story. The order of the testing methods was also randomized,

⁸ Whether the participant can influence the tempo/timing of the 'unfolding' of the pictures

but, for organizational reasons, unmoderated testing was either the first or the last. The moderated tests were conducted and recorded via Zoom. In this way, each session consisted of a meeting with an experimenter via Zoom, where the person told two stories with the help of two testing methods, and of the autonomous testing (beforehand or afterward) via *soscisurvey.de*. In the Zoom meeting, we ensured that the participant did not tell both stories directly – one after the other – and that other tests distracted the persons between the two narrations (these tests included several tasks on working memory, inhibition, word recognition etc.). One Zoom session lasted about 45 minutes; the narration of the stories was the first and the last task.

After completing all three narrations with all three methods, we asked the participant to fill out a final questionnaire with questions about how they felt during the testing and which of the testing methods they liked better.

6.3 Results

I evaluated the use of the different methods from three perspectives: I analyzed the narratives on the macrostructural level (according to the MAIN protocol) (Section 6.3.1) with regard to their similarity of the performance between the three testing methods, the evaluation of the different testing methods by the participants (based on data from the final questionnaire, Section 6.3.2), and by the experimenter (Section 6.3.3). For this last perspective, I interviewed all three experimenters to determine which of the two moderated testing methods worked best in their perceptions. The analysis includes a total of 30 narrative.

6.3.1 Results: Macrostructure

Regarding the perspective of the data, I evaluated the scores for *story structure* (following the MAIN protocol, Max=17 points, for more details, see Gagarina, Klop et al., 2019). Additionally, to assess the *story complexity*, I counted the percentage of verbalized Attempt/Outcome sequences (AO), Goals only (G), or Goal/Attempt/Outcome sequences (GAO).

Table 3 shows the results for story structure, percentage AO-sequences, percentage single Gs and percentage GAOs for all narratives and by method of testing, (PPP, Link, and unmoderated), and by story (Baby Bird = BB, Baby Goat = BG, and Cat). Regarding the three testing methods, the differences between them are marginal. The slightly lower scores for the PPP are an artifact of another outcome, which relates to the narratives of the Baby Goats story. Some participants reinterpreted the story's beginning and said that the baby goat was swimming or playing in the water, not drowning and that the mother goat wanted the baby to come out of the water. This verbalizing of the episode results in a lower score for the respective narratives.⁹ As we had slightly more Baby Goat stories told using the PPP method, this explains the difference between the methods and the comparatively lower score and lower percentage of

⁹ Similar references to a reinterpretation of the Baby Goat story are found in Antonijević et al. (2022, 10), who attributes this to the size of the pictures. They assume that the participant did not see the pictures clearly or large enough, which led to the reinterpretation. That, of course, may be, but it would not explain why this affects the Baby Goat narratives exclusively.

verbalized GAO-sequences in the case of the Baby Goat stories. When all three stories told by one individual were analyzed, no variation due to the different testing methods was found. On the contrary, most people (6) told all three stories in all three methods identically down to the details of structural complexity.

In sum, I found no differences in telling the stories due to the different testing methods. Hence, there were no differences regarding the macrostructure of the elicited narratives.

Table 3: Results: Macrostructure.

	Story Structure (Max = 17)	AO (%)	G (%)	GAO (%)
Total (N = 30)	13.9	17.8	3.3	78.9
PPP (N = 10)	13.7	16.7	6.6	76.7
Link (N = 10)	13.9	16.7	3.3	80.0
Unmoderated (N = 10)	14.0	20.0	0.0	80.0
BB (N = 10)	14.3	16.7	0.0	83.3
BG (N = 10)	13.0	16.7	10.0	73.3
Cat (N = 10)	14.3	20.0	0.0	80.0

6.3.2 Participant: Assessment

To gain insight into the evaluation of the test procedures by the participants, they were asked to fill out a questionnaire, in which they should compare and rate the settings. In addition, I looked at how the participants coped with the testing situation at the technical level. All participants coped well and mastered the technical issues, except for two cases of unmoderated testing, where some technical problems appeared that led to empty audio files and therefore had to be repeated.

When asked which setting was most natural for telling a story, the participants indicated that both moderated methods were equally good and felt natural, but the unmoderated one was worse and more unnatural and unknown. Another question concerned the non-shared knowledge (i.e., in which setting was it best imaginable that the addressee did not see the pictures). In this case, the link variant was named as the best solution, followed by the unmoderated variant. Participants did not mention the PPP variant here. In a final question, the participants were asked to give their overall assessment: Which method was found most pleasant overall. The Link variant received the most positive assessment, followed by the PPP variant. The unmoderated variant received the worst ratings. However, the assessments were somewhat better when the unmoderated test was the final task.

6.3.3 Experimenters: Assessment

To include the experimenters' perspectives, I asked the three experimenters (all trained students with experiences in collecting data) to rate and reflect on the test situations. In general, they agreed that the PPP variant is technically easy to handle. It needs no further explanation as the

experimenter starts and controls the story. This outcome is in line with the statement of the straightforwardness of the PPP testing method in Antonijević et al. (2002, p. 9). However, they pointed out that when using PPP, they often had the feeling of interrupting the flow of the story by asking if they could go ahead with the next pair of pictures. This problem did not occur with the Link variant, but the experimenters stated that this method required a little more time and explanation at the beginning of the test. Initially, they had to explain to the participants how to open the link and move to the next page. Nevertheless, after this introduction, the storytelling flowed smoothly. Regarding unmoderated testing, the most crucial point for the experimenters is that they do not need to be present, so using this method saves time. At the same time, there is a higher level of uncertainty whether everything works as it should, and if not, one can lose the data.

6.4 Summary and conclusion: Pilot study

The pilot study results show that the choice of the testing method does not impact the macrostructure of the resulting narratives; no differences due to the test design were found. This means that the decision for a specific testing method may depend on other factors. As pointed out, such other factors can be the degree of non-shared knowledge and the participant's autonomy in unfolding the pictures. The need for autonomy must be considered from multiple perspective (in some cases, it may be desirable that the experimenter has more control over the procedure); nevertheless, it is a relevant factor in the case of adults. Both factors are not met in the PPP method. The unmoderated method ensures non-shared knowledge and autonomy. At the same time, this testing saves the experimenter's time, but enhances the danger of technical failures. It is also the most unnatural setting for telling a story. Nevertheless, persons with high digital literacy, who are already aware of the testing conditions coped well with this situation. Assuming participants are already familiar with eliciting picture-based stories in general (e.g., because they have previously told a MAIN story in a moderated method) and who have the appropriate digital literacy, this method seems to be well suited for use. The moderated method via Link also guarantees non-shared knowledge and autonomy. At the same time, testing via Link enables the most natural setting for telling a story: there is an interaction with a real person. However, the narrator directs the action himself/herself and is not interrupted or influenced by the experimenter. In cases of technical problems or questions the present experimenter can help, but otherwise he/she does not have to interfere. For these reasons, I generally consider the moderated method via Link as the best choice.

Finally, I want to hint at the limitations of the study. I obtained data from a small sample of young adults with a high level of digital literacy. The results are not generalizable and not transferable to older adults or adults with different backgrounds. Still, they show us what we, in general, must consider before testing MAIN with adults.

7 Conclusion

This paper has provided an overview of some necessary considerations for adapting the MAIN for use with an adult population by considering the similarities and differences of the overall

contexts of testing with adults compared to children, on the one hand, and in different conditions (in-person vs. remote), on the other. I outlined the process of adapting the instructions for testing MAIN with adults and considered some follow-up questions for using them in a remote context. To investigate these questions, I carried out a pilot study with 10 young German-speaking adults. The results of the pilot study showed the successful transfer from in-person testing to three different remote testing methods: two moderated and one unmoderated. The participants told three MAIN stories (Baby Goats, Baby Birds, and Cat) each in another testing method. The analyses of the resulting 30 narratives showed that the testing methods did not influence the macrostructure expressed by the participants. However, the participant and the experimenters experienced the three methods differently. I concluded that I recommend the so-called Link method for remote testing with adults, but that under certain conditions, the unmoderated method (when testing a second story and with participants with a high level of digital literacy) or the PPP method (especially for participants without corresponding digital literacies) are also applicable.

This paper has covered the general background for testing MAIN with adults in a remote context. It has answered some crucial questions and showed that the transfer to an adult and a remote context has been successful. At the same time, some questions remained open, such as the transfer to testing with older persons or those with cognitive impairments or the highly relevant question of addressing the story (addressee). Hence, there is a need for further studies.

References

- adult*, adj. And n. (n.d.). In *Oxford English Dictionary Online*. Oxford University Press. Retrieved 17 February 2023, from <https://www.oed.com/view/Entry/2821>
- Andreoletti, C., Veratti, B. W., & Lachman, M. E. (2006). Age differences in the relationship between anxiety and recall. *Aging & Mental Health*, *10*(3), 265–271. <https://doi.org/10.1080/13607860500409773>
- Antonijević, S., Collieran, S., Kerr, C., & Ní Mhíocháin, T. (2022). Online assessment of narrative macrostructure in adult Irish-English multilinguals. *Frontiers in Psychology*, *13*. <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.916214>
- Bauer, J.-R., Martinez, J. E., Roe, M. A., & Church, J. A. (2017). Consistent Performance Differences between Children and Adults Despite Manipulation of Cue-Target Variables. *Frontiers in Psychology*, *8*. <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01304>
- Beaman, K. V., & Buchstaller, I. (2021). *Language Variation and Language Change Across the Lifespan*. Routledge.
- De Bot, K., & Schrauf, R. W. (2009). *Language development over the lifespan*. Routledge.
- Fiani, R., Henry, G., & Prévost, P. (2022). Macrostructure in narratives produced by Lebanese Arabic-French bilingual children: Developmental trends and links with language dominance, exposure to narratives and lexical skills. *Linguistic Approaches to Bilingualism*, *12*(4), 446–478. <https://doi.org/10.1075/lab.20035.fia>
- Gagarina, N., & Bohnacker, U. (2022a). A new perspective on referentiality in elicited narratives: Introduction to the Special Issue. *First Language*, *42*(2), 171–190. <https://doi.org/10.1177/01427237211062300>

- Gagarina, N., & Bohnacker, U. (2022b). Storytelling in bilingual children: How organization of narratives is (not) affected by linguistic skills and environmental factors: *Linguistic Approaches to Bilingualism*, 12(4), 407–412. <https://doi.org/10.1075/lab.00035.int>
- Gagarina, N., Bohnacker, U., & Lindgren, J. (2019). Macrostructural organization of adults' oral narrative texts. *ZAS Papers in Linguistics*, 62, 190–208.
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., & Walters, J. (2019). MAIN: Multilingual Assessment Instrument for Narratives – Revised. *ZAS Papers in Linguistics*, 63, 1–36.
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics*, 56, 1–140.
- Gerstenberg, A., & Voeste, A. (2015). *Language development: The lifespan perspective*. John Benjamins.
- Hannes, A. (2021). *Untersuchung narrativer Fähigkeiten monolingualer deutschsprachiger Erwachsener mit einer digitalen Adaption von MAIN (Multilingual Assessment Instrument for Narratives)*. Unpublished BA Thesis, Humboldt University of Berlin.
- He, S. (2021). Applying Stepped Task in Remote Unmoderated Test: A Case Report. In C. Stephanidis, M. Antona, & S. Ntoa (Eds.), *HCI International 2021 – Posters. HCII 2021. Communications in Computer and Information Science*, 1419. Springer. https://doi.org/10.1007/978-3-030-78635-9_5
- Hertzum, M., Borlund, P., & Kristoffersen, K. B. (2015). What do thinking-aloud participants say? A comparison of moderated and unmoderated usability sessions. *International Journal of Human-Computer Interaction*, 31(9), 557–570. <https://doi.org/10.1080/10447318.2015.1065691>
- Hržica, G., & Kuvač Kraljević, J. (2022). Referential choice in stories with characters of one or different genders: A study of monolingual Croatian speakers. *First Language*, 42(2), 216–233. <https://doi.org/10.1177/0142723720949728>
- Jažić, I., Gagarina, N., & Perovic, A. (2023). Case marking is different in monolingual and heritage Bosnian in digitally elicited oral texts. *Frontiers in Psychology*, 13. <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.832831>
- Karay, Y., Schaubert, S.K., Stosch, C., Schüttpeitz-Brauns, K. (2015). Computer versus paper--does it make any difference in test performance? *Teach Learn Med.*, 27(1), 57–62. <https://doi.org/10.1080/10401334.2014.979175>
- Karl, K.B. (2021). *Instructions for (online) MAIN testing with adults*. Unpublished presentation held on March 26th 2021.
- Karl, K.B. (in prep.). *Narratives across the lifespan and in different languages: An empirical study on narratives from Russian, Polish and German speakers in different stages of life and with varying addressee*.
- Krasnoshchekova, S., & Kashleva, K. (2019). Narrative Competence of Adult L2 Russian Learners. *Journal of Psycholinguistic Research*, 48(3), 617–641. <https://doi.org/10.1007/s10936-018-9622-3>
- Leeson, H.V. (2006). The Mode Effect: A Literature Review of Human and Technological Issues. *Computerized Testing, International Journal of Testing*, 6(1) 1–24, https://doi.org/10.1207/s15327574ijt0601_1
- Lindgren, J., Tselekidou, F., & Gagarina, N., (this volume). Acquisition of narrative macrostructure: A comprehensive overview of results from the Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics*, 65, 111–132.

- Mattiauda, E., Hassiotis, A., & Perovic, A. (2022). Narrative language abilities in adults with Down syndrome: A remote online elicitation study using the Multilingual Assessment Instrument for Narratives (MAIN). *Frontiers in Communication*, 7. <https://doi.org/10.3389/fcomm.2022.841543>
- McClelland, T., Cuevas, J. (2020). A comparison of Computer Based Testing and Paper and Pencil Testing in Mathematics Assessment. *The Online Journal of New Horizons in Education*, 10(2), 78–89.
- Nalbantoglu Yilmaz, F. (2021). Comparing Paper-Pencil and Computer-Based Tests: A Meta-Analysis in The Sample of Turkey. *Eurasian Journal of Educational Research* 93, 279–300. <https://doi.org/10.14689/ejer.2021.93.13>
- Nascimbeni, F., & Vosloo, S. (2019). *Digital literacy for children: Exploring definitions and frameworks*. Scoping Paper. UNICEF Office of Global Insight and Policy.
- Pfeffer, K. (2015). *Selektieren, verknüpfen, sprachlich umsetzen: Zu viel für Kinder mit einer Sprachentwicklungsstörung? Narrative Fähigkeiten bei Kindern mit und ohne Sprachentwicklungsstörung im Grundschulalter*. Idstein.
- Pratt, A. S., Anaya, J. B., Ramos, M. N., Pham, G., Muñoz M., Bedore, L. M., & Peña E. D. (2022). From a Distance: Comparison of In-Person and Virtual Assessments With Adult–Child Dyads From Linguistically Diverse Backgrounds. *Language, Speech, and Hearing Services in Schools*, 53(2), 360–375. https://doi.org/10.1044/2021_LSHSS-21-00070
- Punch, S. (2002) Research with Children: The Same or Different from Research with Adults? *Childhood*, 9, 321–341. <https://doi.org/10.1177/0907568202009003005>
- Remote Online Language Assessment: Eliciting Discourse from Children and Adults | Frontiers Research Topic*. (n.d.). Retrieved 17 February 2023, from <https://www.frontiersin.org/research-topics/21524/remote-online-language-assessment-eliciting-discourse-from-children-and-adults#articles>
- Roch, M., & Hržica, G. (2020). The role of receptive vocabulary and sentence comprehension: Narrative comprehension by Croatian-Italian bilingual children 5–7 years old. In U. Bohnacker & N. Gagarina (Eds.), *Developing Narrative Comprehension: Multilingual Assessment Instrument for Narratives* (pp. 171–196). John Benjamins. <https://doi.org/10.1075/sibil.61.06roc>
- Rodina, Y. (2020). Using LITMUS-MAIN with Norwegian-Russian bilingual children growing up in Norway. *ZAS Papers in Linguistics*, 64, 163–168. <https://doi.org/10.21248/zaspil.64.2020.570>
- Relawati, A., Primanda, Y., Zamroni, G.M. (2022). Unmoderated Remote Usability Testing: An Approach during Covid-19 Pandemic. *International Journal of Advanced Computer Science and Applications*, 13(1), 283–289.
- Schrauf, R. W. (2009). Longitudinal designs in studies of multilingualism. In K. De Bot, K & R. W. Schrauf (Eds.), *Language Development Over the Lifespan* (pp. 245–270). Routledge.
- Sheng, L., Shi, H., Wang, D., Hao, Y., & Zheng, L. (2020). Narrative Production in Mandarin-Speaking Children: Effects of Language Ability and Elicitation Method. *Journal of Speech, Language, and Hearing Research*, 63(3), 774–792. https://doi.org/10.1044/2019_JSLHR-19-00087
- Smolinsky, L., Marx, B. D., Olafsson, G., & Ma, Y. A. (2020). Computer-based and paper-and-pencil tests: A study in calculus for STEM majors. *Journal of Educational Computing Research*, 58(7), 1256–1278. <https://doi.org/10.1177/0735633120930235>
- Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology*, 84, 1041–1053. <https://doi.org/10.1037/0022-3514.84.5.1041>

- Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *Discourse processing: Multidisciplinary perspectives* (pp. 53–120). Ablex.
- Vogels, J., & Lindgren, J. (2022). The Development of Referring Expression Use from Age 4 to 7 in Swedish-Speaking Children. *Discourse Processes*, 59(9), 722–744. <https://doi.org/10.1080/0163853X.2022.2132794>
- Westby, C. E. (2005). Assessing and facilitating text comprehension problems. In H. Catts & A. Kamhi (Eds.), *Language and reading disabilities* (pp. 157–232). Allyn & Bacon.
- Whitbourne, S. K. (1976). Test Anxiety in Elderly and Young Adults. *The International Journal of Aging and Human Development*, 7(3), 201–210. <https://doi.org/10.2190/8HA7-Q3LE-3VVT-P227>