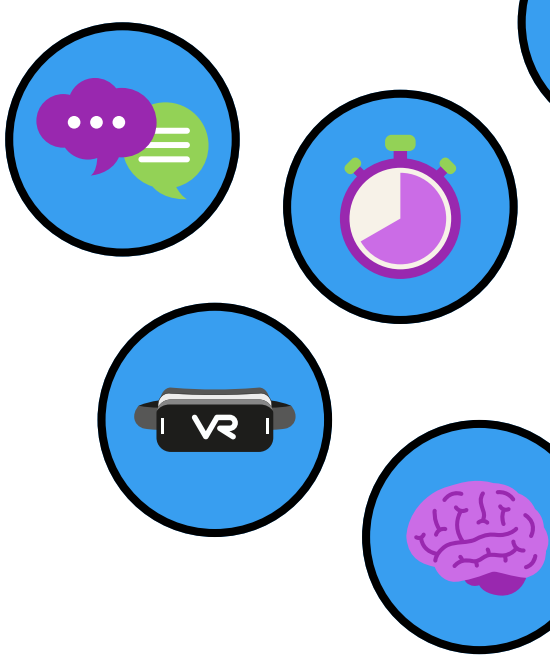


The Effect of Feedback Timing and Flow on Young Professionals' Voice Usage in Virtual Reality (VR) Public Speaking Training.

Emma Lodder (s2836513)

e.lodder@student.utwente.nl

Department of Educational Science and Technology University of Twente

Key words: Immediate Feedback, Delayed Feedback, Flow, Public Speaking, Virtual Reality

Supervisors:

Dr. B. J. Kollöffel.

(b.j.kolloffel@utwente.nl)

Ilona Friso-van den Bos

(i.friso-vandenbos@utwente.nl)

Abstract

Providing young professionals with formative feedback during public speaking training is crucial for improving public speaking skills. Formative feedback can be distinguished based on timing: immediate or delayed. Existing studies revealed inconsistencies in the effect of immediate feedback, revealing it could both positively and negatively impact performance enhancement during public speaking training. This research explored the effect of feedback timing on voice usage of young professionals and the mediating factor of flow experience in a Virtual Reality (VR) public speaking context. Building on prior research of flow theory, this study explored whether immediate feedback acts as a disruptive stimulus to the flow experience, thereby potentially hindering voice usage development. A quasi-experimental, repeated measures design was employed to measure voice usage within two conditions. Participants presented three times lasting three minutes in a VR application called Ovation, in which one condition received immediate feedback, and the other condition received delayed feedback. A voice use rubric and a Psychological Flow Scale (PFS) questionnaire were applied to investigate the flow experience and voice usage. Data was analyzed by executing repeated measures ANOVA, ANCOVA, a linear regression analysis, and a Sobel test. Participants showed an increase in overall voice usage scores over time, regardless of receiving immediate or delayed feedback. Delayed feedback had a more positive impact on the average voice usage over time. Participants receiving immediate feedback revealed higher scores on subcategories "filler words" and "appropriate use of inserted pauses" compared to the delayed feedback condition during the second presentation. The flow experience score of participants receiving immediate feedback was significantly lower compared to the delayed feedback condition. There was no mediation effect of flow experience in the relationship between immediate feedback and voice usage.

Keywords: Immediate feedback, delayed feedback, flow, public speaking, virtual reality.

Table of Contents

Introduction

Improving public speaking skills can benefit young professionals, as these skills are essential for effective performance across various workplace environments (Dunbar et al., 2006; Smith & Sodano, 2011). Providing young professionals with formative feedback during public speaking training can be a crucial factor in skill enhancement, as highlighted by several studies (e.g., De Grez & Leuven, 2009; Kerby & Romine, 2009; van Ginkel et al., 2015, 2017). In public speaking training contexts, formative feedback often involves specific and personalized feedback on the presenter's performance, as shown by Belboukhaddaoui and van Ginkel (2019) and van Ginkel et al. (2017). Van Ginkel et al. (2015) developed design principles to support public speaking competence, emphasizing the importance of feedback and its appropriate timing, whether provided immediately or delayed. However, previous research such as Delucenay et al. (2017) and Dihoff et al. (2004) revealed inconsistent findings on the effect of delayed versus immediate feedback on performance. Moreover, delayed feedback mostly results in positive learning effects, while immediate feedback could positively and negatively impact learning (Shute, 2008).

Research on the role of immediate versus delayed feedback for performance enhancement in public speaking contexts seems limited and shows discrepancies. Specifically, a study by Tanveer et al. (2015) showed that immediate feedback helped improve the presenter's variation in speaking rate. Conversely, Chollet et al. (2015) revealed that immediate feedback, compared to other forms of feedback, is less effective as it distracts the speakers. A more recent study by Belboukhaddaoui and van Ginkel et al. (2019), revealed no significant differences between delayed and immediate feedback impacting the presenter's performance but recommended similar follow-up studies due to the limitations of their study. These contrasting findings underscore the need for further investigation into how immediate versus delayed feedback affects public speaking skills, particularly given concerns that immediate feedback might distract and hinder skill enhancement.

One possible explanation for the negative impact of immediate feedback is a disrupted flow experience. Flow is an experience in which a person is completely immersed in a task (Csikszentmihalyi, 1975; Kang et al. 2022; Volante et al. 2018). The theory's originator, Csikszentmihalyi (1990), relates flow with the "optimal experience" as it is related to a strong feeling of excitement and enjoyment. Furthermore, Schüler (2007) and Engeser and Rheinberg (2008) found a positive relationship between the flow experience and students' performance. Conversely, research by Keller and Blomann (2008) and Keller and Bless (2008) revealed no relationship between flow and performance when participants played computer games. In addition to contrasting findings in existing research, the effect of the flow experience on public speaking skills remains uninvestigated.

The flow experience is characterized by nine components representing experiential states, according to Engeser et al. (2012). These components include "absorption", "effortless control" and

“intrinsic reward” (Norsworthy et al., 2023). Thus, factors that might hinder the flow experience interfere with these core components during the execution of an activity. For instance, the flow experience can be interrupted by self-awareness and performance evaluation during the task (Csikszentmihalyi & Bennett, 1971). Hence, immediate feedback may hinder experiential states of flow and affect performance, although this has yet to be investigated.

In contrast to factors that might hinder flow, technologies inducing flow experiences have been researched, acknowledging Virtual Reality (VR) as a potential contributor (Triberti et al., 2021). Particularly, VR technology can offer opportunities for skill development and action, it can reveal insights into behavior through personalized feedback, and the individual can maintain a sense of control while interacting with the tool (Gaggioli et al., 2003). Moreover, Kang et al. (2022) identified flow as a mediating factor between VR media characteristics and learning transfer, although their context differed, and their research did not address feedback timing. In addition to being a potential contributor to flow, VR can be a useful tool for training individuals’ presentation skills (Kothgassner et al., 2012), allowing them to present for a diverse and large audience and receive detailed feedback (van Ginkel et al., 2019). The effectiveness of VR public speaking applications requires ongoing evaluation due to technological innovations (Poeschl, 2017). Furthermore, Narciss (2008) states that implementing sufficient feedback strategies in technological learning environments remains difficult, and most providers fail as they rely on instinctive approaches rather than proven design principles. Moreover, van Ginkel et al. (2019) emphasize the need for comprehensive research into the effectiveness of feedback types in VR public speaking training.

Consequently, the appropriate timing of formative feedback can be essential in improving public speaking skills. However, research comparing the effects of immediate and delayed feedback on public speaking skills remains scarce and reveals inconsistent findings. While delayed feedback appears to result in positive performance outcomes, immediate feedback can positively enhance performance but also act as a distraction for the speaker. One possible explanation is that immediate feedback negatively influences the flow experience, which in turn affects performance. However, the relationship between immediate feedback, flow, and performance has not yet been explored. Therefore, this research investigates how immediate versus delayed feedback influences voice usage, while also assessing the mediating role of flow experience in a VR public speaking tool. The focus on voice usage can be justified as voice usage is a key component of public speaking competence (Mulder, 2014) and feedback on voice usage can be derived in VR as revealed by researchers (e.g., Belboukhaddaoui & van Ginkel, 2019; Chollet et al., 2022). This research contributes to identifying effective or ineffective formative feedback strategies for improving voice usage in VR training and extends flow theory and research by evaluating its application and impact in the context of public speaking.

Theoretical Framework

Formative Feedback

In public speaking training, formative feedback is acknowledged as a potentially crucial factor in improving presentation skills, as shown by various research (e.g., De Grez & Leuven, 2009; Kerby & Romine, 2009; van Ginkel et al., 2015, 2017). Specifically, formative feedback can affect the learning process of complex behavior, such as developing public speaking competence (De Grez et al., 2009; van Ginkel et al., 2017). Formative feedback can be defined as information provided to the learner aiming to change the behavior or thinking of a learner, improving learning, and, as a result, helping them to achieve a specific goal (Shute, 2008). The opposite, summative feedback, seems to be concentrated on the final judgment of the learner's performance (Mcalpine, 2004). For the current study, the focus is on task-level formative feedback. According to Shute (2008), this includes particular and timely feedback on the learner's performance on a task instead of summarized feedback, including more general feedback provided to multiple learners. This decision can be justified as feedback in public speaking training contexts, often involves specific and personalized feedback on the presenter's performance (e.g., Belboukhaddaoui & van Ginkel, 2019; Chollet et al., 2015, 2022; King et al., 2000; Tanveer et al., 2015; van Ginkel et al., 2017).

The key goal of formative feedback is to expand the learners' skills, understanding, and knowledge within a certain context (Shute, 2008). Moreover, formative feedback can have several benefits when used in public speaking. For instance, research by Kerby & Romine (2009) showed that presenters could better understand the strengths and weaknesses of their presentation and how they could improve it because of feedback. In addition, research by Tanveer et al. (2015) showed that feedback helped improve the presenter's variation in speaking rate. Chollet et al. (2015) revealed that feedback could improve public speaking skills such as using pauses appropriately, intonation, and gestures.

In public speaking training, formative feedback can be provided from different sources. To be specific, several researchers (e.g., De Grez et al., 2009b; Tsang, 2018; van Ginkel et al., 2017) describe that feedback derived in public speaking often involves self-assessment, teacher, or peer feedback. Van Ginkel et al. (2017) showed that feedback provided by peers or a teacher positively impacted the presenting behavior and attitudes of a learner. Conversely, the effect of self-assessment remained limited. Self-assessment may involve different approaches. For instance, the presenter could reflect on one's performance by watching a recorded video and answering specific questions (van Ginkel et al., 2017) or keeping a reflective logbook to evaluate one's performance (Tsang, 2018). Furthermore, new technologies expand the possibilities of providing feedback through digital learning environments (Narciss, 2008), such as video reflection systems (Bower et al., 2011) and Virtual Reality

(Belboukhaddaoui & van Ginkel, 2019; van Ginkel et al., 2019). According to Tsang (2018), sufficient feedback implementation is crucial to assist the individual in self-assessment. Yet, according to Narciss (2008), implementing sufficient feedback strategies in technological learning environments is difficult, and most providers fail as they rely on instinctive approaches rather than proven design principles.

Feedback Timing

As mentioned, formative feedback includes feedback that is delivered timely and can be provided by different sources to improve performance. However, a distinction in timing can be made between deriving feedback immediately or delayed. Immediate feedback is provided directly after the learner has responded to a specific element during a task, and delayed feedback can be delivered in several minutes, hours, or an extended period after the task is completed (Shute, 2008). Van Ginkel et al. (2015) developed design principles to support public speaking competence, emphasizing the importance of feedback and its appropriate timing, whether provided immediately or delayed. Specifically, immediate feedback can affect performance by allowing the presenter to learn and adapt repeatedly during the presentation (King et al., 2000). While presenting, the speaker often concentrates more on their intended message, paying less attention to managing unconscious aspects of their communication, such as eye contact and vocal pace. Thus, immediate feedback can nudge the speaker, aiming at improving these presentation delivery aspects (King et al., 2000). Additionally, immediate feedback can correct certain behaviors early, which can result in efficient retention of the desired behavior (Phye & Bender, 1989).

Delayed feedback benefits the learner by allowing them to practice for the next performance based on the suggestions, which can support long-term memory retention (King et al., 2000; Narciss, 2008). Several researchers named this the “delayed retention effect” (e.g., Phye & Andre, 1989; Schroth & Lund, 1993). Regarding this effect, Schroth and Lund (1993) showed that feedback with a shorter delay had less memory retention effect compared to feedback with a longer delay. This could happen because when a person receives feedback immediately and close to task execution, their memory of the wrong response could hinder developing and memorizing new accurate responses (Kulik & Kulik, 1988). In addition to its benefits, delayed feedback does not interrupt the learning process on how to respond correctly, which immediate feedback might do, but rather reduces the frequency of errors over time (King et al., 2000).

While the timing of feedback seems to play a crucial role in developing public speaking competence, research (e.g., Delucenay et al., 2017; Dihoff et al., 2004; Metcalfe et al., 2009) revealed inconsistent findings on the effect of delayed versus immediate feedback on performance. According to Shute (2008), delayed feedback mostly results in positive learning effects, while immediate

feedback can positively and negatively impact learning. For instance, a learner might rely on immediate feedback in other situations when it is unavailable, and therefore, a decrease in mindful behavior can occur. Furthermore, for problem-solving activities, Corno and Snow (1986, as cited in Shute, 2008) showed that immediate feedback can be interruptive for a learner, which can negatively influence the learning process. Even though public speaking might not directly involve problem-solving, it does involve multiple cognitive, behavioral, and affective processes, according to Blunck (1997), Bower et al. (2011), and Morreale et al. (1993). Thus, the negative effect of immediate feedback could be present in public speaking. Moreover, research by Chollet et al. (2015) revealed that immediate feedback distracted presenters and suggested further investigation into the effects of immediate feedback. In addition, a more recent study by Belboukhaddaoui and van Ginkel (2019) recommended follow-up studies on the effect of immediate feedback on public speaking performance, including researching the impact of delayed feedback.

In sum, formative feedback seems crucial for improving public speaking skills such as gestures, speaking rate, inserting pauses, and intonation. Choosing the appropriate timing for deriving feedback, either immediate or delayed, could impact public speaking skills; however, studies revealed contradictory results on the effect of immediate feedback. Immediate feedback may have a negative impact on performance due to its disruptive nature. Specifically, it could interrupt the experience of flow which could be crucial in optimal performance.

The Experience of Flow

Csikszentmihalyi (1975) defined flow as "the holistic sensation present when we act with total involvement" (p. 43). Building on this, more recent studies such as Kang et al. (2022) and Volante et al. (2018) conceptualized the flow experience as complete immersion in a task, acknowledging Csikszentmihalyi, as the theory's originator. Csikszentmihalyi's initial work (Csikszentmihalyi & Bennett, 1971) primarily associated flow with playful experiences, such as games. However, in his later publications (e.g., Csikszentmihalyi, 1990; Moneta & Csikszentmihalyi, 1996) the role of flow in the learning context was also emphasized. Even though Csikszentmihalyi conceptualized flow more than 40 years ago, there seems to be a general agreement on the definition of flow and its core components (Engeser et al., 2012). Flow, as conceptualized by Csikszentmihalyi (1975), includes six components that describe subjective experiential states. These components form the flow experience and are summarized in the following paragraph.

Starting with the component "merging of action and awareness" defined as a state where the individual is conscious of their actions but not the underlying awareness of those actions. Secondly, the "centering of attention" involves the individual being highly concentrated on the task. Accordingly, distracting stimuli should be avoided to ensure that someone is focused on their task to

achieve the main goal. “Loss of self-consciousness” means that the individual is not aware of the success rate of their performance nor having self-reflective thoughts. Consequently, flow seems to be disrupted by self-awareness and performance evaluation during the task. In addition, when the individual feels one can handle the demands within the environment and one’s actions, “the feeling of control” arises. “Coherent and non-contradictory demands” involves clear goals and unambiguous feedback, meaning that the individual can identify when the activity is executed successfully and understands the structure of a task. Lastly, “autotelic nature” relates to the experience being intrinsically rewarding, meaning that the individual is satisfied by the experience itself. Thus, the individual does not rely on external motivational factors, such as rewards or goals.

In addition to the original components two other components were later introduced: “distortion of temporal experience of time” and “challenge-skill balance” (Engeser et al., 2012). Specifically, when experiencing flow, an individual is completely involved in the activity and can forget the time or get a distorted sense of time (Csikszentmihalyi, 1990, 1996) this refers to the “distortion of temporal experience of time”. Several researchers (e.g., Pearce et al., 2005; Volante et al., 2018) reported a “distorted sense of time” as a key dimension of the flow experience in their studies. Further, the role of the balance between the challenges of the environment and skills of the individual is noted in research (e.g., Csikszentmihalyi, 1996; Moneta & Csikszentmihalyi, 1996; Schüler, 2007). For instance, Schüler (2007) found that participants experienced a stronger flow if the difficulty of a task was balanced with the participant's skills. An imbalance of this aspect can lead to anxiety and worry if the demands in the environment are too challenging to meet the skills of the individual (Moneta, 2021; Barthelmäs & Keller, 2020). Conversely, if the skills exceed the demands in the environment, it can lead to boredom. Besides two additional components, later studies split the original component “coherent, non-contradictory demands” into two components, “clear goals” and “unambiguous feedback” according to Engeser et al. (2012). As a result, the flow experience is understood to consist of nine fundamental components.

While researchers agree on the global definition of flow, it seems more difficult to decide upon the measurement instrument for flow (Moneta, 2021; Engeser et al., 2012). For instance, research reveals inconsistencies in defining the essential components for measuring the flow experience (Norsworthy et al., 2023). Work by Engeser et al. (2012) and Heutte et al. (2021) discussed inconsistencies and critiques of the nine-component model’s construct validity, Swann et al. (2018) revealed conflicts with alternative psychological theories, and Peifer and Tan (2021) and Norsworthy et al. (2021) explored how studies differ in selecting measurement methods for flow. Furthermore, researchers such as Barthelmäs and Keller (2020) and Norsworthy et al. (2023) argued that certain components, such as a perceived fit of skills and demands, function not as a component but as a pre-condition of flow. Additionally, based on a thorough review of 230 flow-related research by

Norsworthy et al. (2021), only half of the nine components might be recognized by learners in educational contexts. The review revealed that flow researchers across different domains and contexts commonly conceptualize three core components (absorption, effortless control, and intrinsic reward) that specify the flow experience, and other original dimensions often function as antecedents of flow or outcome themes. Moreover, existing flow measurement instruments mostly assess none, a few, or one of the core components of the flow experience (Norsworthy et al., 2021). Therefore, Norsworthy et al. (2023) recently developed and tested an instrument that seems to capture these core experiential states. Consequently, although the original nine components are fundamental to understand the flow theory, the current study focuses on measuring the three core components using the instrument of Norsworthy et al. (2023) to capture the essence of flow and reduce the risk of potential construct validity concerns. As discussed, these components are absorption, effortless control, and intrinsic reward. "Absorption" refers to a state in which awareness and action merge, characterized by high focus and attention on the task. "Effortless control" can be defined as the feeling of control and a sense of fluidity in one's action, and "intrinsic reward" involves experiencing high levels of positive emotions, such as enjoyment, during flow.

As mentioned, formative feedback is crucial to improve public speaking competence; however, experiencing flow can also lead to positive performance outcomes. Csikszentmihalyi (1990) relates the state of flow with the "optimal experience," as the flow experience is connected to a strong feeling of excitement and enjoyment. Multiple studies revealed the positive impact of flow on the individual. In particular, the study of Schüler (2007), revealed a positive effect of flow on performance. Specifically, their research showed a positive relationship between students experiencing flow during learning activities and their final grades. Additionally, the study of Engeser and Rheinberg (2008) found a positive relationship between participants experiencing flow during exam preparation and performance at the semester's end. Conversely, research such as Keller and Blomann (2008) and Keller and Bless (2008) found no effect of flow on performance when participants played specific computer games. Moreover, a potential relationship between flow and performance might only exist if the individual perceives the activity as important (Engeser and Rheinberg, 2008). Specifically, Engeser and Rheinberg (2008) found that in tasks with no crucial consequences and low perceived importance, a balance between skills and challenges can lead to flow. Conversely, when the stakes are high, flow is more likely to occur if the skills exceed the challenges. As a result, a higher level of skills and, thus, expertise can lead to better performance (Engeser and Rheinberg, 2008). Additionally, as far as current research indicates, the effect of flow on public speaking skills remains uninvestigated.

In sum, the flow experience can happen when an individual becomes completely immersed in an activity. While this experience could potentially occur during public speaking, it has yet to be

investigated, and the effect of flow on public speaking skills remains unexplored. Besides the crucial role of formative feedback, experiencing flow might also contribute to performance. Conversely, factors that might hinder flow may include factors that interferes with the core components (experiential states) during the execution of a task. As previously mentioned, immediate feedback could negatively impact public speaking training, potentially hindering flow and, consequently, impacting performance.

In contrast to factors that might hinder flow, technologies inducing flow experiences have been researched, acknowledging Virtual Reality (VR) as a potential contributor (Triberti et al., 2021). According to Gaggioli et al. (2003), this can be explained by the characteristics of VR that might stimulate flow. For instance, VR can offer opportunities for skill development and action, reveal insights on behavior through personalized feedback, and the individual can maintain a sense of control while interacting with the tool. On top of that, research (e.g., Belboukhaddaoui & van Ginkel, 2019; Chollet et al., 2022; Poeschl, 2017; Siddiqui et al., 2023; van Ginkel et al., 2019) showed that VR can be a practical tool for the individual to practice public speaking skills.

Virtual Reality (VR) training

Virtual Reality (VR) is an experience in which a person is totally immersed in a virtual environment through projected displays or using head-mounted devices (Feng et al., 2018; Oyelere et al., 2020). It operates by stimulating multiple senses such as vision, sound, and touch, thus creating an illusion of being present in the virtual world (Mandal, 2013; Oyelere et al., 2020). VR finds application across various domains such as gaming (e.g., Shafer et al., 2019), tourism (e.g., Tussyadiah et al., 2018), military (e.g., Lele, 2013), education (e.g., Schöne et al., 2019), and skills training, including in the public speaking context (e.g., Poeschl, 2017). This study explores VR's application in training presentation skills through head-mounted displays, as investigated by several researchers such as Palmas et al. (2019) and van Ginkel et al. (2019).

As mentioned, Virtual Reality (VR) can be beneficial in training public speaking skills, enabling the individual to practice in front of a diverse and large virtual audience (Poeschl, 2017). Studies, such as those by Slater et al. (2006), indicate that the reaction of a user to a virtual audience can be similar to their responses in real-life settings, affirming VR as an effective training tool. Another benefit is that VR offers a safe learning environment, which can be relevant for someone who is not secure about doing the task in real life (Xie et al., 2021). For instance, VR allows users to train comfortably at home (Xie et al., 2021) and provides a safe environment where mistakes have reduced consequences compared to making these mistakes in real life (Poeschl, 2017). Moreover, technological advancements have transformed VR public speaking training by introducing immediate and delayed feedback functionalities that might help presenters enhance their performance (van Ginkel et al.,

2019). However, it is not yet certain whether there is a distinction in effectiveness between delivering feedback messages immediately or after the performance in VR (Belboukhaddaoui & van Ginkel, 2019).

There seem to be two types of immediate feedback in VR public speaking training. One type involves interactive virtual audiences that implicitly respond in real-time to a speaker's presentation (e.g., Chollet et al., 2015; Chollet et al., 2022), while the other seems to involve the delivery of explicit messages on behavior such as facial expressions or use of the voice, as explored by Tanveer et al. (2015) and Belboukhaddaoui & van Ginkel (2019). This study focuses on explicit, immediate feedback, addressing challenges such as the potential for such feedback to be distracting, as noted by Chollet et al. (2015) and Tanveer et al. (2015). Immediate feedback can be derived through technology measuring the participant performances by analyzing facial expressions and speech patterns, as shown in the research of Tanveer et al. (2015). Research on the effects of immediate feedback within VR public speaking training remains limited.

Delayed feedback can be delivered in multiple ways. In the study of van Ginkel et al. (2019) delayed feedback was provided after the performance in a VR public speaking environment, based on a report on the speakers' performance, explained by an expert. Similarly, delayed feedback can be provided through a personalized after-action report, including tailored advice and graphs depicting eye contact and facial expressions (Chollet et al., 2022). Belboukhaddaoui and van Ginkel (2019) provided delayed feedback to the presenter on eye contact and voice use after their presentation in the form of a feedback message within the VR environment. Thus, it appears that in VR public speaking training contexts, delayed feedback, whether in report or message form, is provided after a presentation on the speaker's performance. The advantages of delayed feedback are notable. For example, van Ginkel et al. (2019) revealed that delayed feedback in VR improved crucial behavioral, cognitive, and attitude aspects of public speaking. In addition, Chollet et al. (2022) found that, despite limited training duration, a personalized after-action report could significantly improve speakers' eye contact and facial expressions compared to implicit real-time feedback from a virtual audience.

As previously stated, VR might also contribute to stimulating the flow experience and, thus, performance. However, the connection between flow and performance within a VR public speaking training context appears uninvestigated. In serious game VR studies, such as those by Volante et al. (2018) and Bodzin et al. (2020), the effect of flow has been researched, including how flow can result in a disrupted perception of time. Further, Kang et al. (2022) explored flow's mediating role between VR media characteristics and learning transfer in semiconductor processes and facility training. Their findings supported flow as a mediating factor, advocating the support of flow in VR applications for knowledge and skills improvement. Nevertheless, VR media characteristics do not include immediate

feedback and the context differs. Therefore, it is still unknown whether the flow experience has a crucial role in improving public speaking skills within a VR public speaking training context.

In brief, VR public speaking training can support the individual to improve public speaking skills. Specifically, formative feedback, derived immediately or delayed, can be provided to the presenter in VR public speaking training, offering detailed insights into the speaker's performance. Moreover, VR could be a potential factor in simulating the experience of flow and, thus, performance. As mentioned, voice usage is a common public speaking skill that can benefit from feedback in VR public speaking training. Training presentation skills, including voice usage, is considered important, as it is a key component of communication competence (Mulder, 2014). Furthermore, possessing public speaking skills can be beneficial in numerous contexts, ranging from professional to personal settings (Chollet et al., 2022; Tsang, 2017). In the next paragraph the public speaking skill voice usage is explored further.

Voice Usage

Public speaking skills are beneficial for young professionals and are commonly used and needed in the workplace or for job interviews (Tsang, 2017). Effectively speaking in public requires an individual to possess several verbal and non-verbal abilities, such as setting a clear goal for the presentation, employing effective transitions, selecting words carefully, using different speaking rates, pitch, and intensity, and showing nonverbal behavior that reinforces the spoken explanations (National Communication Association, 1998; Bower et al., 2011). Not all abilities can be addressed in this study. Therefore, this study investigates "voice usage" during VR public speaking training. This focus can be justified as existing research demonstrates that voice usage in VR environments can be measured, and feedback can be given (van Ginkel et al., 2019). Additionally, research (e.g., Chollet et al., 2015, 2022; Tanveer et al., 2015) showed that immediate and delayed feedback could both impact voice use aspects. Moreover, voice usage seems to cover different aspects, including intonation, articulation, speaking rate, volume, pauses, and filler words as revealed by van Ginkel et al. (2019), making it a suitably complex and measurable competence for this study. Each aspect of voice usage will be explained below.

According to the literature, "intonation" is part of the "prosodic" elements of a speech (House, 2006; De Baer & Feryn, 2020), in which timing, the quality of a voice, and rhythm are beneficial (House, 2006). Intonation, a prosodic element, can be defined as movements or variations of voice tones (Cho & Dewaele, 2021; De Baer & Feryn, 2020). This includes the tone or pitch level of the voice going up and down (inflections) and the placement of these variations. Moreover, the placement of the pitch's variation is important as it helps a speaker to structure certain information for the listener and clarifies whether the speaker is asking a question or emphasizing information

(Elbert & Dijkstra, 2014; House, 2006). For instance, in the sentence, 'skincare *can* cause serious skin damage' ('can' is emphasized by a higher tone level compared to the other words), the high possibility of the damage is emphasized. Further, the intonation in a sentence rises if the speaker wants to ask a question (De Baer & Feryn, 2020). Specific intonation used by the speaker will define if the listener can interpret the intended message, therefore, (poor) intonation can determine the success rate of the communication (House, 2006). Besides clarifying the intended message, fluctuations in pitch can also express the positive or negative emotion behind the speech (Cho & Dewaele, 2021). For instance, research by Weger et al. (2007) showed that positive emotional speeches tend to include more frequent and upward inflections, while negative emotional speeches tend to be related to less frequent inflections and more downward inflections. Concretely, these insights shed light on the importance of using intonation deliberately during public speaking.

In addition to intonation, "articulation" is another crucial element of voice usage skill. It seems that articulation refers to words being clearly and understandably spoken, which, according to Si Na et al. (2020) contributes to the ability to communicate and share ideas effectively every day. Studies on public speaking (e.g., Nishimura & Hashida, 2018; Si Na et al., 2020; Siddiqui et al., 2023) indicate that when words are spoken understandable, this aspect is defined as "pronunciation". Opening the mouth correctly during speaking can contribute to speech delivery, such as pronunciation, as well as nonverbal delivery (Nishimura & Hashida, 2018). For instance, if a recipient cannot hear what a speaker says, they can guess the words by observing the mouth shape of the presenter. Levelt (1989) wrote an overarching book on articulation research to provide a theoretical framework for this voice aspect and the mental information processes connected to it. According to the author "articulation flows automatically, at a rate of about fifteen speech sounds per second, while we are attending only to the ideas we want to get across to our interlocutors" (P. 13). It is executing the "phonetic plan" by using specific muscles within the neck and mouth. Concrete, the phonetic plan includes a mental plan of how the words should be spoken before articulating them.

Another element of voice usage includes "speaking rate". Guyer et al. (2021) and Montes et al. (2019) refer to speaking rate as the words used per minute. According to Montes et al. (2019) and Toastmasters International (2011), an effective speaking rate in English includes approximately 120 to 160 words per minute. Research on speaking rates reveals that a variation of low or high speaking rates can contribute to the persuasion of the presenter. Jiang and Pell (2014), (2017) and Scherer and Wolf (1973) found that presenters who talk confidently increase their speaking rate, while unconfident or doubtful presenters have a lower speaking rate. Yet, a high speaking rate could result in recipients being unable to process the intended message, decreasing the perceived confidence of the presenter and the presentation's quality (Guyer & Fabrigar, 2017; Moore et al., 1986). Furthermore, an extremely high speaking rate can weaken the quality of a strong argument but can

strengthen a weak argument, as revealed by Guyer and Fabrigar (2017). Consequently, when intending to make a strong statement, a lower speaking rate might increase the quality and impact of the argument.

Using intelligible “volume” during a speech is another aspect of voice usage. Volume is a critical element of a speech, as the speaker can make themselves clear and understandable (Okrasa et al., 2022). Furthermore, Okrasa et al. (2022), realized thresholds to categorize volume levels as “quiet”, “moderately quiet”, “acceptable”, “moderately loud”, and “too loud”. Knowlton and Larkin (2006) found that “acceptable” loudness is a precondition for recipients to take advantage of other voice usage aspects, such as intonation. In addition, research (Jiang & Pell, 2017; Kimble & Seidel, 1991; Scherer & Wolf, 1973; Van Zant & Berger, 2020) revealed that confident presenters talk with louder volumes compared to unconfident presenters. For instance, Jiang and Pell (2017) investigated how speakers use their voices and how these were perceived by recipients. The study revealed that speakers assigned to perform a confident speech spoke louder, and recipients assessed them as more confident afterward.

The final voice usage element addressed in this study is the correct use of silent “pauses”. A distinction can be made between long and short pauses. Long pauses should be inserted at the paragraph’s ending or when a speaker changes the topic (Sokoreva & Shevchenko, 2022), and short pauses happen at the end of an utterance (Wennerstrom & Siegel, 2003). After making a key statement, a longer pause can also be used to leave a stronger impression on the audience, a technique frequently employed by representatives and public speakers according to Sokoreva and Shevchenko (2022). An audience can acknowledge a presenter as reliable, potent, and entertaining through eye contact and gestures, as well as by inserting silent pauses (Montes et al., 2019). Moreover, pausing is crucial to breath and plan what to say next (Sokoreva & Shevchenko, 2022). Besides unfilled pauses, they can also be filled with words such as “like”, “um” or “uh”, which can distract and annoy recipients (Rhodes & Frandsen, 1975). Laske and DiGennaro Reed (2024) and Montes et al. (2019) refer to filler words as a “speech disfluency”, and, therefore, should be avoided. According to Laske & DiGennaro Reed (2024) fillers include two categories “filler words” such as “you know” “like” and “so” and secondly “filler sounds” such “um” and “ah”. Fillers are mostly used when the speaker is unable to speak the intended words within a speech, indicating the speaker is searching for words on what to say next (Clark & Fox Tree, 2002).

Research Questions

Formative feedback is crucial for improving complex behavior, such as public speaking competence. However, integrating sufficient feedback strategies in innovative technologies such as VR, seems to be a difficult task. A critical decision lies in whether formative feedback is derived immediately, delayed, or both; however, whether timing yields consistent outcomes on performance remains questionable and requires further investigation. Immediate feedback could have a negative effect on the speaker. Nevertheless, explanations for this negative effect do not seem to exceed the definition of “distraction”.

A possible explanation for immediate feedback not being as effective as delayed feedback could be due to a disruptive experience of flow. While Kang et al. (2022) found a mediating effect of flow on learning transfer, Csikszentmihalyi (1990) Schüler (2007) Engeser and Rheinberg (2008) claim that flow positively affects performance, and Triberti et al. (2021) discuss that VR contributes to the experience of flow, evidence of the effect of flow on public speaking skills in VR training remains limited.

Understanding the effect of immediate versus delayed feedback on flow and voice usage is scientifically relevant, as it can contribute to identifying effective or ineffective strategies for improving public speaking skills with emerging technologies and extends flow research by evaluating its application in VR training environments in the context of public speaking. Hence, this study investigates the relationship between the timing of feedback, the experience of flow, and voice usage.

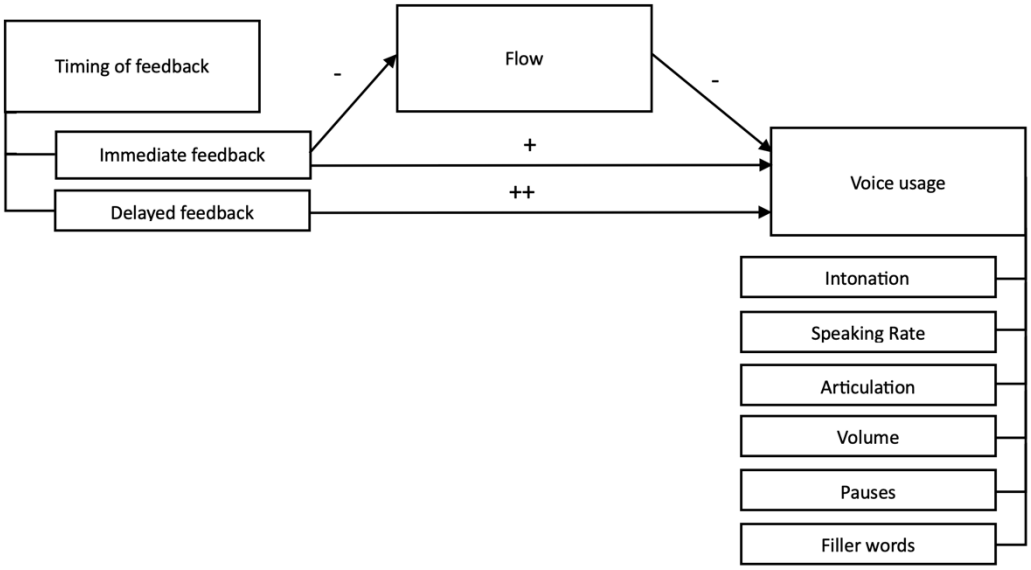
Based on the theoretical framework, the research questions of this study are:

1. “How does immediate feedback, compared to delayed feedback, during Virtual Reality (VR) public speaking training influence voice usage among young professionals?”
2. “How does immediate, compared to delayed feedback, during Virtual Reality (VR) public speaking training influence the flow experience? “
3. “How does the flow experience influence voice usage during Virtual Reality (VR) public speaking training, and to what extent does it mediate the effect between immediate feedback and voice usage?”

A summary of the expected findings is depicted in figure 1.

Figure 1

Expected findings between timing of feedback, flow, and voice usage.



Stemming from previous research and theory, the following hypotheses are formulated:

H1. Immediate feedback has a weaker positive effect on voice usage within VR public speaking training compared to delayed feedback.

H2. Immediate feedback negatively affects the flow experience of the speaker within VR public speaking training compared to delayed feedback.

H3. The flow experience positively affects the public speaking skill voice usage.

Method

Research design

In this research, it is investigated if immediate versus delayed feedback (independent variables) affects voice usage (dependent variable) of young professionals within VR, and if flow experience is a mediating factor between immediate feedback and voice usage. This study employed quantitative research, a quasi-experimental repeated measures design where the dependent variable (voice usage) and the mediating factor (flow) were measured at three moments in time within two distinct conditions. Participants were randomly assigned to one of the experiential conditions. In one condition, participants received immediate feedback during the second presentation, therefore, this condition is referred to as the "immediate feedback condition". Participants of the other condition received delayed feedback after the second presentation and will be referred to as the "delayed feedback condition". To investigate if the flow experience can mediate the relationship between immediate feedback and voice usage, after each presentation, the participants of both conditions filled in a questionnaire. The dependent variable voice usage was measured using a voice usage rubric.

Participants

A convenience sampling method was adopted to recruit participants for the experiments. The sample was obtained through an invitation post on LinkedIn and e-mailing trainee groups within the organization the researcher works for. Within this global aviation organization of approximately 25,000 employees, the trainees are mostly highly qualified young professionals who focus on development and reflection as they work toward finding a fitting position within the organization. Additionally, the trainees fulfill temporary roles in different departments till the traineeship's ending.

Participants signed up voluntarily by contacting the researcher through the invitation. They had to be between 20 and 35 years old and required to give presentations regularly within their jobs or in their future careers to fit the criteria of the sample. The final sample of this research consisted of Dutch and English-speaking young professionals (n=42). Of 42 participants, 22 were randomly assigned to the immediate feedback condition and 20 to the delayed feedback condition.

Procedure

The data collection started after approval from the Ethics Committee of the Faculty of Behavioral, Management and Social Sciences of the University of Twente on 2024-03-13. The corresponding review number is 240097. Participants were scheduled for the experiments via email. Hereafter, preparation instructions were sent, including the instruction to prepare a three-minute presentation with slides. Before the experiment, the slides of the participants were configured into Ovations' environment, and the room was set up with VR equipment (hardware and software), an

iPad for the participants to complete the PFS questionnaire, and a laptop to record the session and stream the Ovation content.

In both conditions, the experiment lasted 30 to 45 minutes depending on the technology and involved the researcher and one participant. Before the experiment started, participants completed an active online consent form, indicating their willingness to participate in this research. The experiment proceeded only after participants gave their consent by signing the form. Participants were informed that they could withdraw from the study at any time without providing a reason. After this, participants received instructions on how to use the hardware, to navigate within Ovation, and on the experimental set-up: participants had to deliver the same three three-minute presentations to a virtual audience in Ovation that they had prepared in advance. It was explained that the researcher would stop their speech after three minutes, regardless of whether it was finished. The researcher also stopped the presentation if the participant was finished before three minutes. After the instructions, the participant put on the head-mounted display and held the controllers, see Figure 2 for an illustration of the procedure. Once the participant indicated their readiness to begin, they clicked on “start speech”. After a participant clicked on “start speech” during each phase in the experiment, the researcher set a timer and started the recording simultaneously. Hereafter, the data collection unfolded as follows:

The initial presentation served as a baseline measurement for both conditions. After three minutes the researcher instructed participants to remove the head-mounted display and complete the PFS questionnaire. After this, participants of the immediate feedback condition received additional instructions before they started their second presentation. Specifically, participants were informed that they would receive immediate feedback in the form of messages based on their performance during their second presentation. They were instructed to focus on feedback on voice usage and adapt it to improve their performance. If participants of both conditions indicated their readiness to perform the second presentation, they again put on the head-mounted display, held the controllers, and clicked on “start speech”.

During the second presentation, the treatment was introduced for the immediate feedback condition. Specifically, immediate feedback was presented during the presentation followed by removing the head-mounted display and completing the PFS questionnaire. Hereafter, the participant put on the head-mounted display held the controllers, and clicked on “start speech” to perform the third presentation.

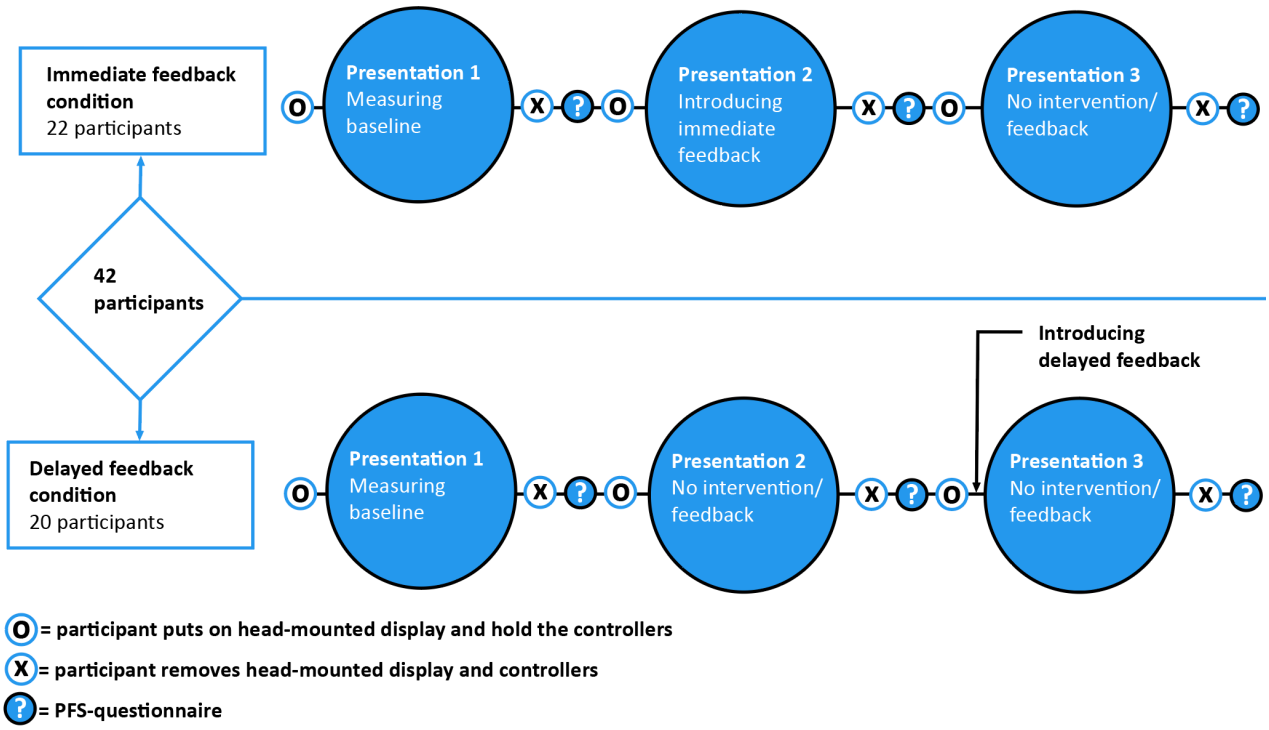
The delayed feedback condition maintained the initial setup and completed the PFS questionnaire after performing their second presentation. At the same time, the researcher set up Ovations’ analytics report, which displayed the results of the participants’ performance of their second presentation. After completing the questionnaire, the participant put on the head-mounted

display, held the controllers, and interacted with the analytics report. The researcher navigated the participant through the analytics report and emphasized feedback regarding voice usage. Before participants of the delayed feedback condition started their third presentation, the researcher instructed the participant to adapt voice usage feedback during their third presentation. Afterward, the participant clicked on "start speech" to perform the third presentation. The third presentation did not include treatments, and afterward, participants in both conditions removed the head-mounted display and completed the flow questionnaire for the final time.

To assess the voice usage of participants, the researcher checked the recordings and evaluated each presentation using the voice usage rubric. To test for interrater reliability, three randomly selected experiments were evaluated by an additional researcher.

Figure 2:

Procedure



Instrumentation

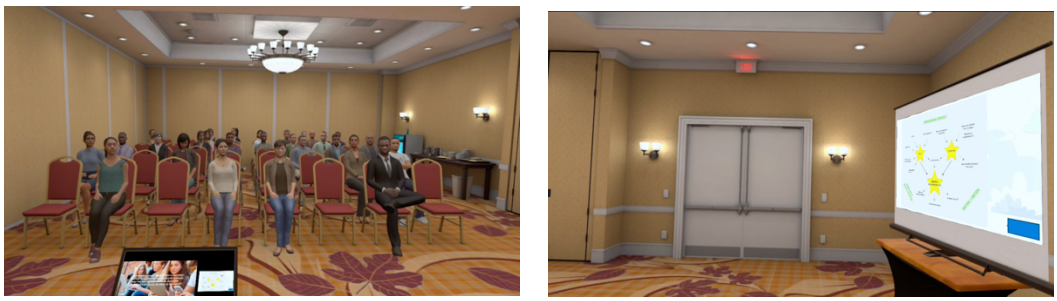
VR hardware and software

The Meta Quest 3 served as the VR hardware of this study. The hardware included a head-mounted display and a pair of controllers. The VR software "Ovation" was used, which enabled the participants to present in a virtual environment to an audience, seamlessly integrated into display projection. Ovation provides different settings in which a user can present, and the audience can be customized, for instance, as a big or small audience or a rudely behaving audience. It is an English-

based software, meaning that the user interface is only available in English; however, it does support 21 languages during a presentation. The Dutch language is not supported, which required the Dutch participants in this study to interact with an English interface and features. For this study, a formal small conference room was chosen, featuring a diverse and casually dressed virtual audience of approximately 30 people who behaved politely. This setting might come close to settings that the participants were used to as they were required to stand in front of the audience and could also use their slides, which were visible both in front of and behind them (see Figure 3).

Figure 3:

Setting of the Formal Small Conference Room with Audience



Note. The images show the perspective of a participant within Ovation. On the left image, the slides of the participant are displayed in the black box. In the right image, the slides are displayed on the beamer behind the participant.

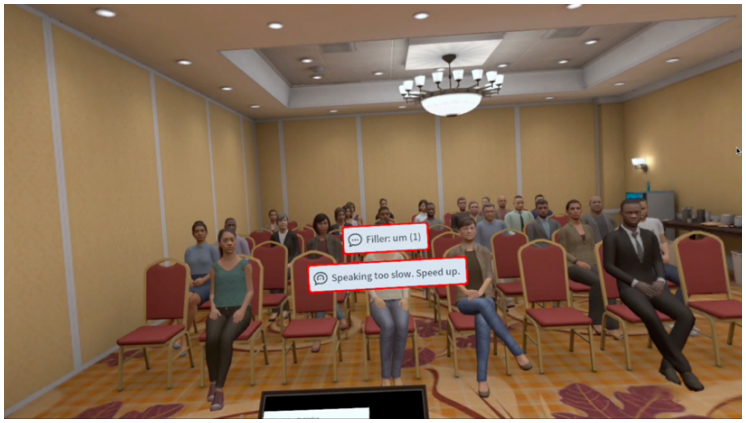
Immediate Feedback

To conduct this research, immediate feedback is a requirement to investigate the impact on the flow experience and voice usage. Therefore, Ovation was configured to provide immediate feedback in English on "voice". For this category, the system generated immediate feedback on four categories: monotony, speaking rate, pauses, and filler words. An immediate feedback message was displayed to a participant when they reached an underperforming threshold according to the software's algorithms. Figure 4 shows an example of immediate feedback messages. For each category, the message features a warning text, an icon, and a red frame emphasizing the alert. The warning text and timing for displaying an immediate feedback message differ by category. More specifically, for the "speaking rate" category, the text "Speaking too slow. Speed up." was displayed when a participant paused too often or had a low speaking rate during their speech. For the "pauses" category, the text warning "Remember to Pause" appeared when a participant infrequently paused during their presentation. "More pitch variation" was displayed during the speech when the participant frequently spoke monotonously, thus using less pitch variation (intonation). Further, if the participant used the filler words "um", "uh" or "yeah" they received a warning. For instance, if a

participant said “um” once during their speech, the warning text included “Filler: um (1)”, indicating the filler word and the number of times it was used. If a participant used the Dutch filler word “ja” (meaning “yes” in English) once, it showed the warning “Filler: Yeah (1)” and, thus, was included in the immediate feedback.

Figure 4:

Immediate Feedback on Voice Usage Displayed in Text Messages



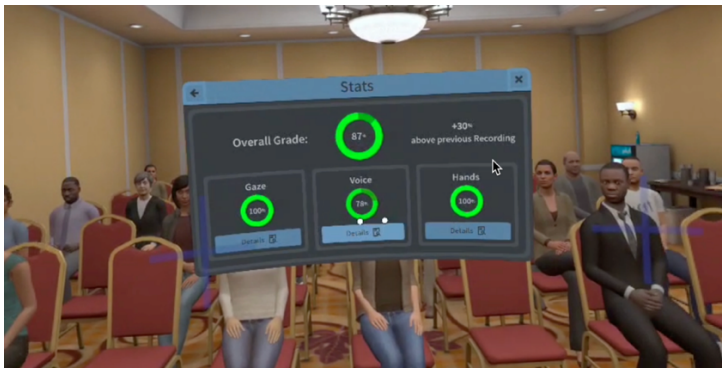
Note. The image shows the perspective of a participant within Ovation. The message above shows what filler word is used and how many times, the message below includes a warning regarding the speaking rate.

Delayed Feedback

The delayed feedback condition received delayed feedback in Ovation for the same category as the immediate feedback condition, namely, “voice”. Delayed feedback was derived as the system analyzed speech patterns and generated an analytics report afterward. First, this report presented an overall summary, including the average performance grade in percentage and detailed information on each category (see Figure 5). Each subcategory included a total score expressed in percentages. Within the report, percentages are mostly framed by a pie chart of color indicating the sufficiency of the performance. For instance, if a participant’s score fell between 100 and 50 percent for a certain category, the pie chart was colored green. If a participant’s score fell between 0 and 50 percent, the color changed to orange; close to 0 percent, it shifted to red.

Figure 5:

Delayed Feedback in the Form of an Analytics Report



Note. The image shows the perspective of a participant within Ovation.

Detailed feedback on "voice" included the same subcategories as the immediate feedback condition. Two graphs were displayed for the subcategory "speaking rate". The "average words per minute" graph displayed the average words used per minute in numbers, plotted against an x-axis where the left side indicated a "slow" and the right side "fast" speaking rate (see Figure 6). The "words per minute over time" graph, displayed time in minutes on the x-axis, and included the frequency of words used on the y-axis.

Figure 6:

Delayed Feedback for the Participant on "Maintain a Comfortable Speaking Rate"

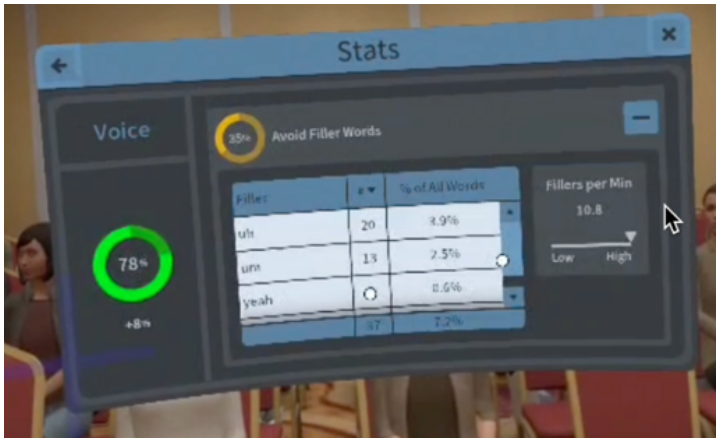
Note. The image shows feedback on speaking rate from a participant's perspective in Ovation.

The subcategory "filler words" included a table and a plot, as displayed in Figure 7. The table included the filler words "uh", "um", and "yeah" if they had been used, along with their frequency and the percentage they represented of all words used during the presentation. The plot showed the

frequency of filler words used per minute plotted on an x-axis where the left side indicated low and the right side a high frequency of filler words used.

Figure 7

Delayed Feedback for the Participant on "Avoid Filler Words"



Note. The image shows feedback on Avoid Filler Words from a participant's perspective in Ovation.

Feedback on subcategory "pauses" included two graphs (see Figure 8). "Pauses per minute" included the frequency of pauses inserted during the speech per minute plotted on an x-axis where the left side indicated low and the right side a high frequency of inserted pauses. The "pause count" graph included a pie chart with color segments representing the duration of pauses in seconds. The numbers within each segment showed the frequency of these inserted pauses.

The subcategory "monotony" included one graph displaying the average pitch variation of the participant, plotted against an x-axis where the left side indicated a "low" and the right side "high" pitch variation.

Figure 8

Delayed Feedback for the Participant on "Use Pauses Regularly"

Note. The image shows feedback on Use Pauses Regularly from a participant's perspective in Ovation.

Voice Usage Rubric

To observe and assess the participant's voice use, an as-yet-unpublished voice use rubric developed by members of the research group of the BMS department of the University of Twente served as a guideline. The original rubric (Appendix B) is employed in similar studies (e.g., Amende, 2024) to assess the voice usage of participants presenting in Ovation and is provided by the supervisor of this study Dr. B. J. Kollöffel. The rubric was tested on interrater reliability in the study of Amende (2024), showing a Cohen's kappa coefficient of 0.83, indicating strong agreement. Some changes were made to the original rubric to adapt it for this study, for instance, some categories were left out and guidelines for scoring with the rubric were added. Figure 9 illustrates a sample of the rubric used for this study, the complete version is available in Appendix A. Concretely, the instrument contains seven criteria, assessing five categories of voice usage: intonation (criteria 1-2), articulation (criterion 3), speed (criteria 4-5), pauses (criteria 6), and filler words (criterion 7). Each criterion is rated with grades 1 to 10, where 1 is absent and 10 is excellent, thus making it an interval-level measurement.

As mentioned above, changes were made to the original rubric. The category "volume" was excluded because presentations were evaluated using video recordings, making it difficult to assess volume explicitly. The original category of "pauses" included two criteria, "inserted pauses" and "appropriate use of pauses" and were merged into one category, "appropriate use of inserted pauses". The decision is based on the understanding, as found in the literature, that effective voice usage is determined by the appropriate placement of pauses and not the frequency of pausing during a speech. Furthermore, each grade has been explained to provide comprehensive guidelines for scoring the categories. It is important to note that these explanations serve as a framework and may not encompass all the criteria within each category. During the evaluation of presentations, these explanations are intended as general guidelines rather than exhaustive descriptions.

Figure 9

Sample of Adjusted Voice Use Rubric

Variation of Intonation

Listen for: changes in pitch throughout the presentation and enthusiasm of the speaker.

1 **Absent** No variation in pitch: the speaker speaks monotonously.	2 **Failure** The variation in pitch is minor, rare, and barely noticeable, with a near-monotone delivery.	3 **Poor** There are very few variations in pitch, with a near-monotone delivery.	4 **Insufficient** Occasional variation in pitch: the speaker attempts to vary tone, but the variation is infrequent and minor.	5 **Mediocre** Moderate variation in pitch: the speaker varies tone inconsistently and speaks slightly more monotonously than engaging.
6 **Sufficient** General variation in pitch: The speaker varies tone sufficiently maintaining a reasonable level of interest, occasionally speaking monotone.	7 **Satisfactory** Good variation in pitch: The speaker's tone variation contributes to the presentation engagement, with infrequent moments of speaking monotonously.	8 **Good** Good variation in pitch: the speaker effectively uses tone changes, with very few moments of speaking monotonously.	9 **Very good** Very good variation in pitch throughout the presentation the speaker's tone changes are highly engaging.	10 **Excellent** There is excellent variation in pitch: the speaker consistently employs dynamic and engaging pitch changes, with no areas needing improvement.

Psychological Flow Scale Questionnaire

While previous research such as Kang (2022) and Volante (2018) measured flow using the flow experience scale constructed by Jackson and Marsh (1996), research by Norsworthy et al. (2023) emphasized troubles among measurement and conceptual aspects of these pre-existing instruments. Therefore, they recently developed an instrument to assess the flow experience among different contexts, which was used for this study. This instrument, an open-access questionnaire, comprises 9 items assessing three dimensions of flow: absorption (items 1-3), effortless control (items 4-6), and intrinsic reward (items 7-9). Each item is rated on a seven-point scale (1=strongly disagree, 7=strongly agree), an interval measurement level, allowing for the computation of the global flow score and subscale scores based on the average responses to the three items within each dimension. In the study by Norsworthy et al. (2023), reliability testing for this instrument resulted in a Cronbach's Alpha coefficient of 0.82, indicating acceptable reliability. Figure 10 illustrates a sample of the questionnaire. The complete version, including the Dutch-translated questionnaire, is available in Appendix C.

Figure 10
Sample Flow Questionnaire

		Strongly disagree	Neutral					Strongly agree
1	I was absorbed in the act/task	1	2	3	4	5	6	7
2	I was highly focused on the task/activity	1	2	3	4	5	6	7
3	All my attention was on the task/activity	1	2	3	4	5	6	7
4	I felt like I could easily control what I was doing	1	2	3	4	5	6	7
5	My actions flowed effortlessly	1	2	3	4	5	6	7

Data Analysis

Data analysis for this study was conducted using SPSS, version 29.0.2.0, and included anonymized quantitative data obtained from the voice use rubric and the PFS questionnaire. Data derived from these instruments were merged into a wide format.

The voice use rubric was tested on interrater reliability by Cohen's kappa, where .81 indicates perfect agreement, and below .60 indicates insufficient agreement (McHugh, 2012). To execute this test, an additional assessor evaluated three randomly chosen experimental sessions. A Cohen's kappa statistic below the acceptable range meant that the scores needed to be revised to achieve higher reliability levels before resuming the statistical analysis. If the Kappa statistic was above .60 the data

collection could continue. Secondly, data derived from the PFS questionnaire was tested for reliability by conducting a factor analysis and deriving Cronbach's alpha coefficient (α). The results of the factor analysis were checked to determine if items were correlated to the respective factor, exceeding the threshold of .32. This threshold was selected as it was used by Norsworthy et al. (2023), indicating poor factor loadings. Afterward, Cronbach's alpha coefficient (α) of every factor was checked for the level of reliability. A coefficient between .7 and .9 indicated that the questionnaire was reliable (Gliem & Gliem, 2003). A repeated measures ANOVA was done to analyze whether feedback timing affects voice usage and if there was a significant difference between the two conditions. First, the 'voice usage' variable was computed by summing up and averaging the scores of the voice usage categories per participant. After, all categories of voice usage and the average voice usage were examined separately to determine if the mean of each variable significantly changed over time and if there was an interaction effect between the immediate and delayed feedback condition. For each ANOVA analysis conducted in this study, first, the assumptions for conducting ANOVA were checked. Secondly, Mauchly's test of sphericity was examined to determine if the sphericity assumption was met ($p > .05$). If so, the sphericity assumed row within the ANOVA table was checked to report the *F* statistic, *p*-value, and effect size. The Greenhouse-Geisser row in the ANOVA table was checked for similar statistics when sphericity was violated. Significance was determined by a *p*-value smaller than .05, and the effect size was measured using partial Eta-squared. The effect size was considered medium between .06 and .14 and large above .14, as noted by Cohen (2013). Additionally, an ANOVA repeated measures was conducted to assess how feedback timing influenced the flow experience and if there was a difference in the flow experience scores between the immediate and delayed feedback condition. The 'flow experience' variable was computed by summing up and averaging the scores of the flow experience dimensions per participant. All dimensions of the flow experience and the average flow experience were examined separately to determine if the mean scores of each variable significantly changed over time and if there was an interaction effect between the immediate and delayed feedback conditions. A one-way repeated measures ANCOVA was conducted to determine if the flow experience was a covariate in the relationship between the feedback timing and voice usage. Additionally, a mediation analysis was done by conducting a linear regression analysis and a Sobel test.

Results

To assess the voice use rubric on interrater reliability, three random presentations including three rounds each were evaluated on seven voice use criteria by a second assessor (corresponding to 63 subjects). This evaluation showed a .39 Cohen's Kappa statistic, meaning there was a fair/moderate agreement (63 subjects, 2 raters, $z = 7.82$, $p < .05$). Together with the second assessor, these scores have been discussed and adjusted. Hereafter, the adjusted scores were used for another interrater reliability test and showed a Kappa statistic of .90, indicating an almost perfect agreement (63 subjects, 2 raters, $z = 16.13$, $p < .01$).

As the Psychological Flow Scale (PFS) Questionnaire of Norsworthy et al. (2023) is a recently developed instrument, a factor analysis was conducted to investigate the relationship between the nine items measuring three dimensions of flow experience. The results of the factor analysis are displayed in Table 1, showing that the items are correlated to the respective factor, exceeding the threshold of .32. Additionally, Cronbach's α coefficient of every factor showed that the level of reliability of each factor is very reliable with "absorption" $\alpha = .84$, "effortless control" $\alpha = .89$, and "intrinsic reward" $\alpha = .89$. After assessing the voice use rubric on interrater reliability and investigating the factor analysis of the PFS questionnaire, the research questions were further analyzed in the section below.

Table 1

Results From a Factor Analysis of a Psychological Flow Scale (PFS) Questionnaire

Item	Factor loading		
	1	2	3
Factor 1: Absorption			
1. I was absorbed in giving the presentation	**0.60**	0.13	0.15
2. I was highly focused on giving the presentation	**1.00**	-0.18	-0.05
3. All my attention was focused on presenting	**0.74**	0.14	0.00
Factor 2: Effortless Control			
4. I felt like I could easily control what I was doing	0.08	**0.83**	-0.20
5. My actions flowed effortlessly	-0.05	**0.89**	0.06
6. There was a sense of fluidity to my actions	-0.11	**0.94**	0.07
Factor 3: Intrinsic Reward			
7. I found the experience rewarding	0.03	-0.10	**0.94**
8. The experience felt satisfying	-0.01	0.07	**0.89**
9. I would like the feeling of that experience again	-0.05	0.06	**0.79**

Note. *N* = 42. The extraction method was exploratory factoring with an oblique (Promax) rotation. Factor loadings above .32 are in bold.

Voice Usage over Time

There are two conditions in this research. In both conditions, participants gave three 3-minute presentations. For the immediate feedback condition, the intervention (immediate feedback) was introduced during the second presentation. The intervention (delayed feedback) for the delayed feedback group was introduced after completing the PFS questionnaire after the second presentation. Table 2 shows the descriptive statistics for the dependent variable "voice usage" and its six categories and the "flow experience" and its three dimensions. To determine "voice usage", the scores of the categories per participant were summed up and averaged. For "flow experience", scores across dimensions were summed and averaged per participant.

To address the first research question, "how does immediate feedback, compared to delayed feedback, during Virtual Reality (VR) public speaking training influence voice usage among young professionals?", a repeated measures ANOVA has been conducted. This test examined if voice usage significantly changed over time and if there was an interaction effect between the immediate and delayed feedback condition. In addition, the effect of immediate versus delayed feedback on the categories of voice usage was tested by conducting the same analysis. Before testing the hypotheses, the assumptions for conducting ANOVA were checked and met. Mauchly's test of sphericity indicated that the sphericity assumption was met ($p = .93$). Therefore, the sphericity assumed row within the ANOVA table was checked to report the results. There is a significant main effect when $p < .05$. The results of the within-subjects test, revealed a significant effect of time on voice usage scores for both conditions with $F(2, 40) = 11.39$, $p < .001$. The effect size, calculated as partial eta-squared (η^2), was .22, indicating a large effect. This indicates that in each condition participants' voice usage scores significantly changed over time. There was also a significant interaction between time and condition $F(2, 40) = 3.81$, $p < .03$, partial $\eta^2 = .09$, indicating that the change of voice usage scores over time differed depending on the participants' condition. These results are visualized in Figure 11, showing that for the immediate feedback condition, voice usage scores remained relatively stable between the second and third presentations. In contrast, voice usage scores increased for the delayed feedback group during the second and third presentations. For both conditions, the voice usage scores in the third presentation improved compared to the first presentation. The in-between-subject effects indicated no significant difference in the voice usage score between the two feedback conditions over time.

Figure 11

Estimated Marginal Means of Voice Usage during the three 3-minute Presentations for the Immediate and Delayed Feedback condition.

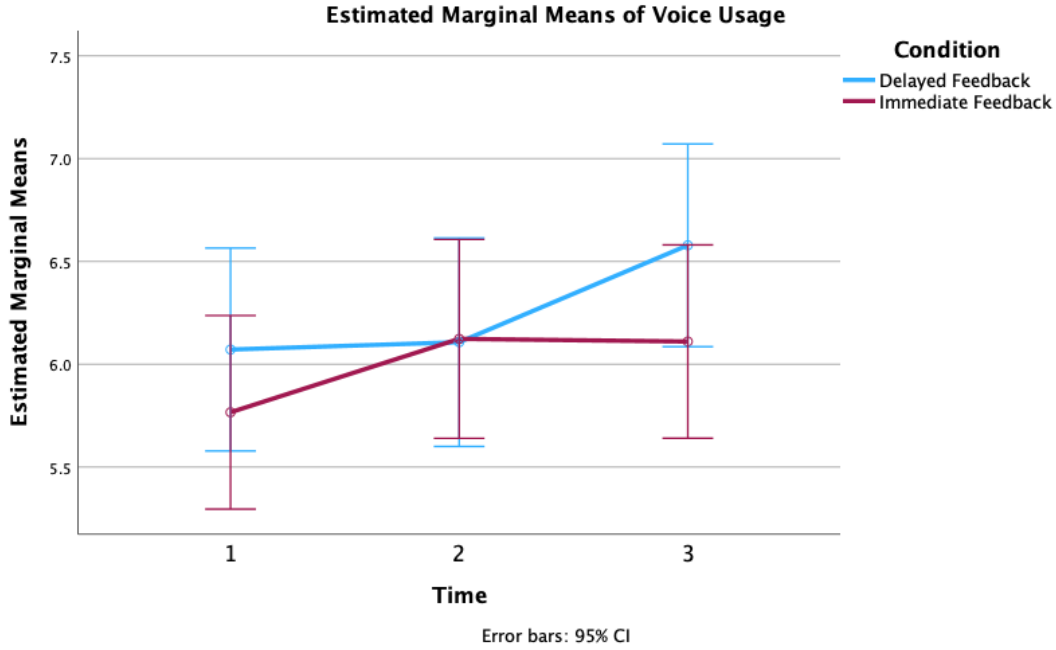


Table 2

Descriptive Results per Variable, Group, and Presentation

Group	Immediate						Delayed					
Variable	*M*1	*SD*1	*M*2	*SD*2	*M*3	*SD*3	*M*1	*SD*1	*M*2	*SD*2	*M*3	*SD*3
Variation of Intonation	6.41	1.60	6.68	1.32	6.50	1.30	6.65	1.87	6.75	1.69	6.70	1.53
Appropriate Use of Intonation	5.68	1.50	5.95	1.40	5.91	1.41	5.85	1.46	6.30	1.49	6.25	1.41
Articulation	7.23	1.60	7.14	1.61	7.18	1.59	7.60	1.57	7.70	1.42	7.65	1.39
Variation in Speed	5.23	1.41	5.23	1.41	5.55	1.06	5.80	1.58	5.75	1.62	5.65	1.39
Appropriate Use of Speed	6.09	1.34	6.14	1.08	6.36	0.95	6.40	1.64	6.55	1.50	6.60	1.27
Pauses	4.95	1.50	5.36	1.36	5.32	1.46	4.80	1.36	4.90	1.62	6.05	1.43
Filler Words	4.77	2.45	6.36	2.08	5.95	2.70	5.40	2.64	4.80	2.69	7.15	2.21
Voice Usage	5.77	1.03	6.12	0.96	6.11	1.11	6.07	1.15	6.11	1.28	6.58	1.07
Absorption	5.45	1.09	4.90	1.60	5.98	0.99	5.63	1.27	5.98	1.20	5.85	1.37
Effortless Control	5.20	1.40	4.30	1.50	5.86	1.02	4.92	1.06	5.43	0.99	5.53	1.27
Intrinsic Reward	5.30	1.20	4.91	1.53	5.91	0.91	5.00	1.32	5.30	1.45	5.73	1.50

Flow Experience	5.31	0.90	4.71	1.31	5.92	0.76	5.18	0.96	5.57	1.09	5.71	1.28

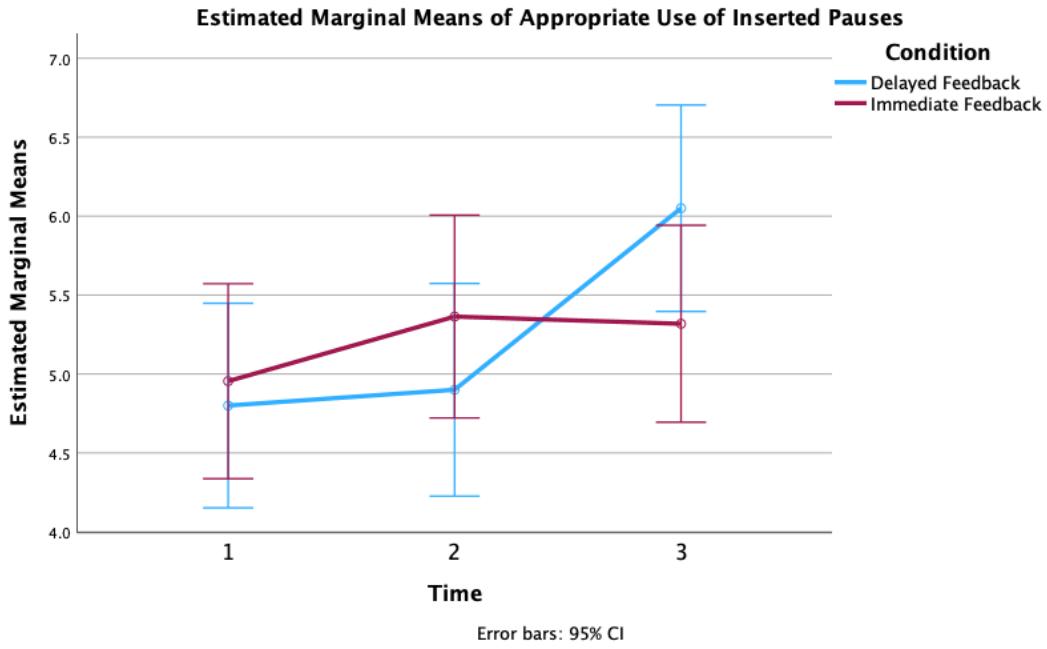
Note. *N* = 42 (*n* = 20 for the delayed feedback group and *n* = 22 for the immediate feedback group). *M*1= mean first presentation; *SD*1 = standard deviation first presentation; *M*2 = mean second presentation; *SD*2 = standard deviation second presentation; *M*3 = mean third presentation; *SD*3 = standard deviation third presentation.

When conducting repeated measures ANOVA to assess the effect on "appropriate use of intonation", Mauchly's test of sphericity indicated that the assumption of sphericity was met ($p = .51$). Therefore, results were drawn from the sphericity assumed row of the ANOVA table. The participants' scores demonstrated significant changes over time for the variable "appropriate use of intonation" ($F(2, 40) = 5.55$, $p < .006$, partial $\eta^2 = .12$). There was no interaction between time and condition, indicating that the change of appropriate use of intonation scores over time did not significantly differ between the two conditions. Additionally, the in-between-subject effects test indicated no significant difference in the appropriate use of intonation score between the two feedback conditions over time.

To assess the effect on "appropriate use of inserted pauses", a repeated measures ANOVA was conducted. Mauchly's test of sphericity indicated that the assumption of sphericity was met ($p = .37$). Therefore, results were drawn from the sphericity assumed row of the ANOVA table. The participants' scores demonstrated significant changes over time for the variable "appropriate use of inserted pauses" with $F(2, 40) = 9.23$, $p < .001$. The effect size, calculated as η^2, was .19, indicating a large effect. There was also a significant interaction between time and condition $F(2, 40) = 5.22$, $p < .007$, partial $\eta^2 = .12$, indicating that the change of the scores of appropriate use of inserted pauses differed over time depending on the participants' condition. These results are displayed in Figure 12, showing that for the immediate feedback condition, the appropriate use of pauses scores increased slightly between the first and second presentation and remained relatively stable between the second and third presentation. In contrast, the scores for delayed feedback between the second and time third presentations increased, displaying a steep incline.

Figure 12

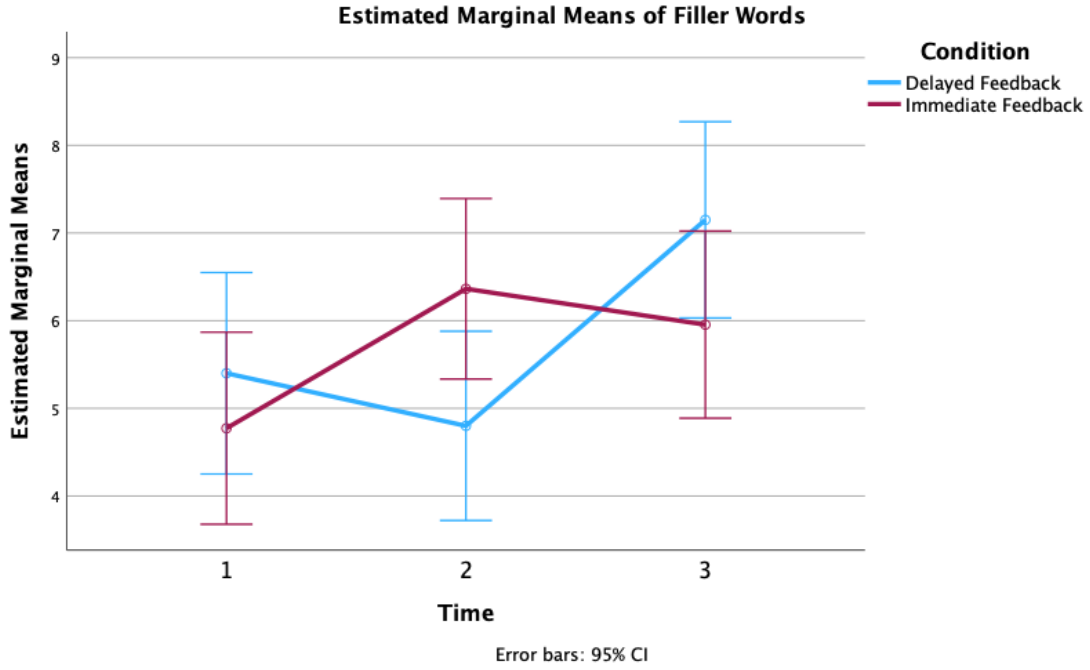
Estimated Marginal Means of Appropriate Use of Inserted Pauses during the three 3-minute Presentations for the Immediate and Delayed Feedback condition.



When conducting repeated measures ANOVA to assess the effect on "filler words", Mauchly's test of sphericity indicated that the assumption of sphericity was met (p = .06). Therefore, results were drawn from the sphericity assumed row of the ANOVA table. Participants' scores of the variable "filler words" revealed significant changes over time with $F(2, 40) = 12.23$, $p < .001$. The effect size, calculated as η^2, was .23, indicating a large effect. There was also a significant interaction between time and condition ($F(2, 40) = 11.68$, $p < .001$). The effect size of $\eta^2 = .23$ indicates a large effect. These results show that the change in the scores of filler words differed over time depending on the participants' condition. Figure 13 displays the change in the use of filler words over time, showing that for the immediate feedback group, the scores of filler words increased from the first to the second presentation, showing a steep positive incline. Hereafter, filler word scores remained relatively stable between the second and third presentations. For the delayed feedback group, the score of filler words slightly decreased between the first and second presentation and increased between the second and third presentation, showing a positive steep incline.

Figure 13

Estimated Marginal Means of Filler Words during the three 3-minute Presentations for the Immediate and Delayed Feedback condition.



Results of the within-subjects test for the corresponding categories of voice usage showed no main effect of time on variation in intonation, articulation, variation in speed, and appropriate use of speed. For these categories, there was also no interaction effect between time and condition.

The Flow Experience over Time

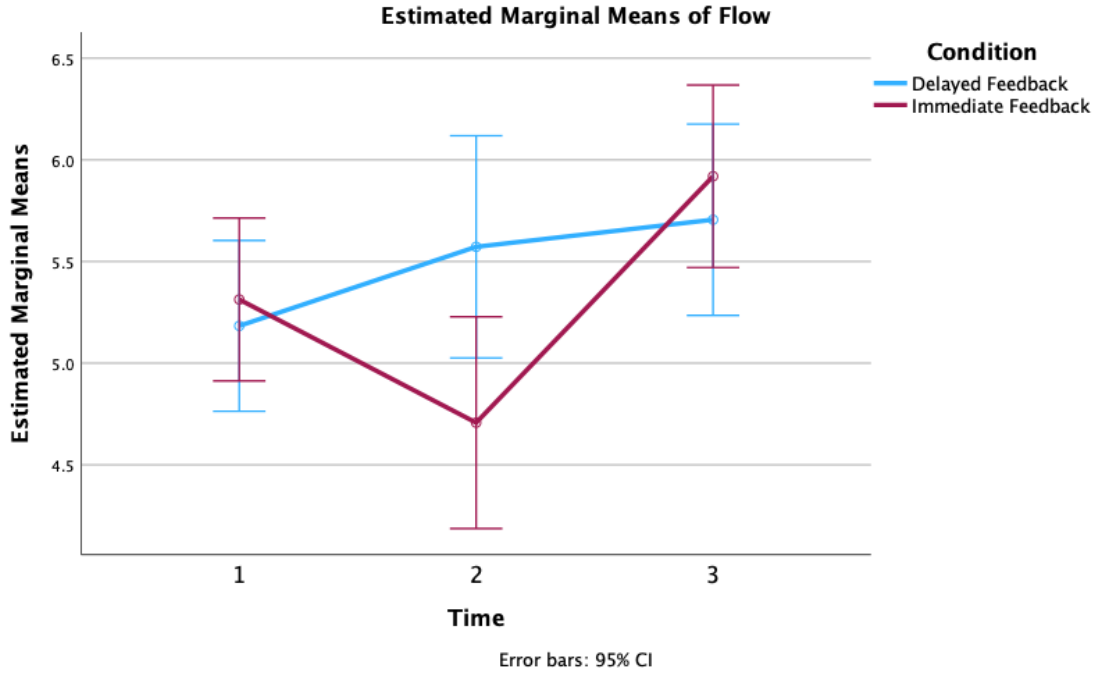
To explore the second question, "how does immediate, compared to delayed feedback, during Virtual Reality (VR) public speaking training influence the flow experience?" a repeated measures ANOVA was conducted to assess whether the flow experience changed over time and if there was an interaction effect between the immediate and delayed feedback conditions. Furthermore, the dimensions of the flow experience were also separately examined by conducting repeated measures ANOVA, to test sub-relationships, as suggested by Norsworthy et al. (2023).

First, Mauchly's test of sphericity was checked, indicating that the sphericity assumption was met (p = .20). Therefore, the sphericity assumed row within the ANOVA table was checked to report the results. Results of the ANOVA test revealed a significant effect of time on the flow experience for both conditions with $F(2, 40) = 12.20$, $p < .001$, indicating that for both conditions, participants' average flow experience score significantly changed over time. The effect size, calculated as partial eta-squared (η^2), was .23, indicating a large effect. There was also a significant interaction between time

and condition $F(2, 40) = 7.42$, $p < .001$. The effect size as partial $\eta^2 = .17$, indicating a large effect. These results indicate that the change of the average flow experience score over time differed significantly depending on the participants' condition. Consequently, for the immediate feedback group, flow experience scores decreased from the first to the second presentation, showing a steep negative incline (see Figure 14). Hereafter, the flow experience score increased between the second and third presentation, displaying a steep positive incline. For the delayed feedback group, the score of flow experience increased between the first and second presentation and increased slightly between the second and third presentation, showing a positive incline.

Figure 14

Estimated Marginal Means of Flow experience during the three 3-minute Presentations for the Immediate and Delayed Feedback condition.



Results of the repeated measures ANOVA test for the dimensions of the flow experience revealed significant changes over time for each variable. For the variable "absorption" the Greenhouse-Geisser row in the ANOVA table was checked, as Mauchly's test revealed that sphericity was not assumed (p .005). Results showed a significant effect of time on absorption for both conditions with $F(2, 40) = 4.04$, $p < .03$, partial $\eta^2 = .09$. There was also a significant interaction between time and condition $F(2, 40) = 6.43$, $p = .005$. The effect size as partial $\eta^2 = .14$, indicating a large effect.

When conducting repeated measures ANOVA to assess the effect on "effortless control", Mauchly's test of sphericity indicated that the assumption of sphericity was met ($p = .82$). Therefore, results were drawn from the assumed row of the ANOVA table. Participants' scores of the variable

"effortless control" revealed significant changes over time with $F(2, 40) = 7.72$, $p < .001$. The effect size, calculated as η^2, was .18, indicating a large effect. There was also a significant interaction between time and condition ($F(2, 40) = 7.72$, $p < .001$). The effect size as partial $\eta^2 = .16$, indicating a large effect.

For the variable "intrinsic reward" the Greenhouse-Geisser row in the ANOVA table was checked, as Mauchly's test revealed that sphericity was not assumed (p .03). Results revealed a significant effect of time on intrinsic reward for both conditions ($F(2, 40) = 18{,}79$, $p < .001$). The effect size as partial $\eta^2 = .32$, which exceeds the threshold of a large effect (.14), indicating that 32% of the variance in the dependent variable can be explained by the effect of time. Additionally, there is a significant interaction effect between time and condition with $F(2, 40) = 3.83$, $p = .03$, partial $\eta^2 = .09$.

The Mediating Role of the Flow Experience

To investigate the third research question, "how does the flow experience influence voice usage during Virtual Reality (VR) public speaking training, and to what extent does it mediate the effect between immediate feedback and voice usage?" first, a one-way repeated measures ANCOVA was conducted to investigate whether the flow experience acts as a covariate in the relationship between feedback timing and voice usage. Specifically, it was examined if the flow experience influenced the effect of feedback timing on voice usage over time.

As the covariate in repeated measures ANCOVA must be independent of the effects of the intervention (immediate and delayed feedback), the participant's flow experience score of the first presentation was taken as a covariate in this analysis. Mauchly's Test of sphericity revealed that sphericity was assumed ($p = .94$). Therefore, the row of sphericity assumed was checked in the ANCOVA table. Results of the within-subjects test revealed no significant effect of time on voice usage across the three presentations for both conditions after controlling for the flow experience ($F(2, 40) = .91$, $p = .41$, partial $\eta^2 = .02$). As previously stated, to explore research question one, the results of the within-subject test ANOVA revealed a significant effect of time on voice usage scores for both conditions ($F(2, 40) = 11.39$, $p < .001$, partial $\eta^2 = .22$). Including the flow experience as a covariate revealed that the flow experience may account for the variability that was contributed to time. Results revealed a significant interaction effect between time and condition after controlling for flow experience, $F(2, 40) = 3{,}82$, $p = .03$, partial $\eta^2 = .09$. Indicating that the change in voice usage over time differs between the conditions, even after controlling for the covariate. The within-subject test showed no interaction effect between time and the flow experience, suggesting that the effect of the flow experience score on voice usage does not significantly differ over time. Lastly, the between-subjects test revealed no significant overall effect of the condition and the flow experience on voice usage after controlling for the covariate.

To examine whether the flow experience accounts for the variability previously attributed to time and to test for a potential mediation effect of the flow experience, a linear regression analysis and a Sobel test were conducted. Specifically, the indirect effect of immediate versus delayed feedback on the flow experience (path a), an indirect effect of the flow experience on voice usage (path b), and an overall and direct effect of immediate versus delayed feedback on voice usage were examined (path c/c'). For the dependent variable, the difference between the voice usage scores of the second and third presentation were computed for each condition. These scores indicate the improvement in voice usage after participants received feedback. The flow experience score of participants in the second presentation was used to determine its role as a mediator. The first step of the analysis showed that the indirect effect of immediate feedback on the flow experience time 2 (path a) was significant (B = -.87, t(39) = -2.32, p = .03). Indicating that immediate feedback predicts a lower flow experience score. The indirect effect of the flow experience on voice usage (path b), was not statistically significant (B = .06, t(39) = .89, p = .38). Additionally, the total effect of immediate feedback on voice usage (path c) was not significant (B = -.16, t(39) = -.93, p = .36), nor was the direct effect of immediate feedback on voice usage with controlling for flow (path c': B = -.12, t(39) = -.66, p = .51). Lastly, a Sobel test was conducted, further confirming that there was no significant mediation effect of flow between immediate feedback and voice usage with z = .58, p = .56.

Discussion

The aim of this research was to investigate how immediate feedback, compared to delayed feedback, during Virtual Reality (VR) training, influenced the public speaking skill voice usage among young professionals. Secondly, it was investigated how the flow experience mediates the effect of immediate feedback during VR training on the public speaking skill voice usage. The conclusions of this study will be explained per hypothesis and research question.

Feedback Timing Effect

Regarding the first research question, "how does immediate feedback, compared to delayed feedback, during Virtual Reality (VR) public speaking training influence voice usage among young professionals?" it was hypothesized that immediate feedback has a weaker positive effect on voice usage compared to delayed feedback. These expectations were based on research indicating that within VR public speaking training, immediate feedback on performance can be more of a distracting factor instead of a supporting factor to participants (Chollet et al. 2015). However, the results from the repeated-measures ANOVA showed that the average voice usage score of participants significantly improved from the first to the third presentation, regardless of receiving immediate or delayed feedback in VR. Nevertheless, the change of the average voice usage score between each presentation differed significantly depending on whether participants received immediate or delayed feedback. Participants who received immediate feedback did not reveal a significant difference in the average voice usage score compared to participants who received delayed feedback. Participants who received delayed feedback after the second presentation scored higher on the average voice usage at the third presentation, compared to participants in the immediate feedback condition, whose average voice usage scores remained relatively stable between the second and third presentations. Thus, delayed feedback has a more positive impact on participants' average voice usage score compared to immediate feedback, confirming the first hypothesis. This is in line with previous studies that showed that immediate feedback could have a positive influence on voice usage during VR public speaking training. For instance, Damian et al. (2015), showed that immediate feedback improved the openness, body energy, and speech rate during public speaking. Additionally, participants in the study of Tanveer et al. (2015) found immediate feedback helpful in enhancing their speaking rate. However, some studies (e.g., Chollet et al., 2015; King et al., 2000) note that immediate feedback can be distracting for participants during their presentation, which might cause a discrepancy in the average voice usage scores between immediate and delayed feedback conditions. Presenters seem to especially favor immediate "sparse" feedback to minimize distraction, rather than feedback continuously presented (Chollet et al., 2015; Tanveer et al., 2015). Depending on the performance, participants of the immediate feedback condition within this study received feedback continuously.

An alternative explanation of a stronger effect of delayed feedback on voice usage could be due to the content included in the feedback. According to Nicol and Macfarlane-Dick (2006), specific strategies need to be considered in sufficient feedback practices to support self-regulated learning and, thus performance. For instance, a good feedback practice clarifies the strengths and weaknesses of performance but also clarifies how to improve it. Tsang (2018) mentioned that feedback such as "the body language was poor or good", does not contribute to these strategies. Furthermore, Tanveer et al. (2015) displayed immediate feedback by mapping the presenters' current performance on a two-point scale of icons, indicating a high or low performance, which resulted in a positive effect on performance. In the current study, participants who received delayed feedback were given an analytics report containing more detailed information on the desired performance, such as a graph including whether the pitch variation (intonation) was high or low and what was desired. Conversely, immediate feedback seems to be misaligned with the strategies, offering mainly a warning of "more pitch variation", instead of providing insights on what performance was desired.

In brief, the weaker positive effect of immediate feedback on voice usage may be due to speaker distraction, as observed in several studies (e.g., Chollet et al., 2015; King et al., 2000), where constant feedback is less favored than sparse feedback (Chollet et al., 2015; Tanveer et al., 2015). Further, feedback can support performance improvement if it contains more descriptive information, such as the expected performance (Nicol & Macfarlane-Dick, 2006). Therefore, the general content included in immediate feedback versus the extensive and detailed content included in delayed feedback may serve as an additional explanation.

In addition to the overall effects of immediate and delayed feedback on the average voice usage score, the categories within voice usage were also investigated for possible sub-relations. While participants who received immediate feedback during the second presentation did not show a significant difference in the average voice usage score compared to delayed feedback, a difference was observed in "filler words" and "appropriate use of inserted pauses" between the two conditions. For instance, participants who received immediate feedback used fewer filler words in the second presentation than in the delayed feedback condition. In contrast, the number of filler words used in the third presentation remained relatively consistent for the immediate feedback condition, while the delayed feedback condition showed less use of filler words. First, the contradiction in findings of the average voice usage and categories of voice usage may be due to the frequency of the kind of immediate feedback presented to participants during the second presentation. To be concrete, the VR technology displayed messages on filler words more frequently to all participants compared to immediate feedback on speaking rate, pauses, and monotony. The reason for participants to score higher on the appropriate use of inserted pauses in the second presentation could be due to a possible correlation with filler words. According to Montes et al. (2019), minimizing filler word usage

may present practical limitations due to the possible correlative relationship with other voice usage elements. Specifically, changes in filler word usage, such as an increase or decrease, can impact other aspects of the voice. For example, using fewer filler words could result in inserting more appropriate pauses. Concretely, the observed increase in just two categories of voice usage may result from the frequent feedback on filler words, which can also influence other voice usage skills, such as the appropriate use of inserted pauses.

As previously mentioned, research (e.g., Damain et al., 2015; Tanveer et al., 2015) revealed that immediate feedback could also positively affect voice usage, which could explain the improvement in filler words and the appropriate use of inserted pauses. King et al. (2000) revealed that immediate feedback especially impacts behavior that requires little attention capacity, such as eye contact. Filler words and appropriate use of inserted pauses may potentially be part of this category. An alternative reason for the other categories being unaffected could be due to cognitive overload. According to the cognitive load theory, people have a maximal capacity for processing information through the working memory system (Schnotz & Kürschner, 2007; Sweller & Chandler, 1991; Van Merriënboer & Sweller, 2010). Furthermore, Miller (1956) states that confusion and errors may occur in someone's output after the maximum capacity to process information is exceeded. Voice usage encompasses multiple categories and, therefore, can be considered a complex skill. Such skills can be described as "*high element-interactive material*", which can induce a high intrinsic cognitive load (Schnotz & Kürschner, 2007; Sweller & Chandler, 1994; Van Merriënboer & Sweller, 2010). In the current study, a high intrinsic cognitive load could have been present as the presenter dealt with multiple aspects simultaneously (e.g., their narrative, voice, gestures, slides, VR). Besides the intrinsic cognitive load inherent to the task, immediate feedback could have caused extraneous load. Within several studies (e.g., Schnotz & Kürschner, 2007; Sweller & Chandler, 1991; Van Merriënboer & Sweller, 2010), it is discussed that a high extraneous load can be generated by learning material and might be the result of presenting too much information that goes beyond the boundaries of the working memory. Therefore, for teaching complex tasks, extraneous load must be minimized to optimize the capacity of the working memory system necessary for learning (i.e., germane load) and to prevent cognitive overload (Schnotz & Kürschner, 2007).

As outlined above, the scores on "filler words" and "appropriate use of inserted pauses" between the second and third presentation remained relatively stable for the immediate feedback condition, while the scores improved significantly for the delayed feedback condition. This reveals that there is a weaker effect of immediate feedback on voice usage when the feedback is removed compared to receiving delayed feedback. A possible explanation may be grounded in research by Corbett and Anderson (2001), who revealed that someone can rely on immediate feedback in situations when it is unavailable. More specifically, participants who received immediate feedback

began the problem-solving activity within a program immediately, whereas removing feedback led participants to wait until finishing the entire program before asking for help. Even though participants of the current study were told that immediate feedback was excluded in the third presentation, it could be the case that participants depended on immediate feedback to improve voice usage aspects.

In sum, the immediate feedback provided during the second presentation may have led to cognitive overload, a concept investigated by several researchers (e.g., Schnotz & Kürschner, 2007; Sweller & Chandler, 1994; Van Merriënboer & Sweller, 2010). As a result, participants might struggled to effectively enhance multiple elements of voice usage, including intonation, articulation, and speaking rate, due to reaching the maximal capacity of processing information. Cognitive overload can affect learning capacity (Schnotz & Kürschner, 2007), which could explain the relatively stable voice usage scores between the second and third presentation. Furthermore, as revealed by Corbett and Anderson (2001), participants may have relied on immediate feedback when it was unavailable, and removal might have limited further voice usage improvement. Additionally, the increase in only two categories of voice usage may be attributed to the frequent feedback on filler words, a skill that can impact other elements of voice usage, such as the appropriate use of inserted pauses. In contrast, delayed feedback appears to enhance all aspects of voice usage simultaneously. This improvement may result from the more extensive and detailed content included in this type of feedback, which, as Nicol and Macfarlane-Dick (2006) note, supports performance enhancement. Lastly, the potential influence of flow experience as an additional explanatory factor is discussed further in this discussion.

Feedback Timing and Flow

Regarding the second research question, “how does immediate, compared to delayed feedback, during Virtual Reality (VR) public speaking training influence the flow experience?” it was expected that immediate feedback negatively affects the flow experience compared to delayed feedback. This was based on the theory of flow (Csikszentmihalyi, 1975) and the framework of Norsworthy et al. (2023), indicating if an individual does not experience a sense of control during a task, becomes distracted, or does not perceive the task as intrinsically rewarding, the intensity of the flow experience can be decreased or disrupted. This hypothesis is supported by the results of this study, revealing that compared to the first presentation, the flow experience scores of participants receiving immediate feedback were significantly lower in the second presentation. Moreover, the flow experience scores increased in the third presentation when feedback was removed. In contrast, the flow experience score of participants in the delayed feedback condition improved significantly during each presentation. The gradual increase in the flow experience scores for participants of the delayed feedback condition can be explained by the core of the revised flow model presented by Barthelmäs and Keller (2020). According to them, the intensity of the flow experience can increase when a person

finds an activity valuable and there is a fit between the environment's demands and the participant's skills. Moreover, if someone engages in a specific activity and experiences flow, they might find the activity more valuable afterward, leading to an increase in the intensity of the flow experience when participating in this activity again (Barthelmäs & Keller, 2020; Engeser & Rheinberg, 2008). Conversely, the intensity of the experience of flow increased significantly for the immediate feedback condition between the second and third presentations. However, this effect may not stem from experiencing flow in the previous presentation but rather because the crucial flow components, such as intrinsic reward, were more strongly experienced in the third presentation when immediate feedback was excluded.

It is also worth considering whether the immediate feedback condition had all the preconditions to induce a flow experience in the second presentation. More specifically, according to Barthelmäs and Keller (2020), a crucial pre-condition is a "perceived fit between skills and challenges". This means that flow is most probably experienced when an individual perceives their skills and challenges of the environment as high (Csikszentmihalyi, 1996; Moneta & Csikszentmihalyi, 1996; Schüler, 2007). Public speaking alone can be a demanding and difficult activity. Moreover, it includes the combination of verbal and non-verbal abilities (Bower et al., 2011; Ortiz et al., 2016) and individuals can be anxious to speak in public (Lee et al., 2002; Tsang, 2020; Wang et al., 2020). Therefore, adding immediate feedback might affect whether the participant perceives their voice usage skills as sufficient, potentially disturbing the balance between the perceived fit between skills and challenges.

To summarize, participants who received immediate feedback reported significant lower flow experiences compared to the delayed feedback condition. However, when immediate feedback was removed, flow levels significantly increased. This suggests that immediate feedback interferes with absorption, effortless control and intrinsic reward. An imbalance of the perceived fit of skills and task demands could explain the increase in the flow experience.

Mediating Effect of Flow

An alternative answer to the first research question could be that a decrease in the state of flow causes participants of the immediate feedback condition to score lower on voice usage compared to participants of the delayed feedback condition. Therefore, another aim of this study was to investigate to what extent the flow experience mediates the effect between the timing of feedback and voice usage. It was hypothesized that the flow experience would positively affect voice usage, and the flow experience would mediate between the immediate feedback and voice usage. Hence, it was expected that immediate feedback would negatively affect the flow experience and, therefore, negatively affect voice usage. In VR public speaking training, the mediating role of flow has not yet

been investigated. However, Kang et al. (2022), found that flow could be a mediating factor and contribute to the application of knowledge and skills within a Virtual Reality semiconductor facility training. Furthermore, a relationship could be expected between flow and performance as the flow experience is intrinsically rewarding and involves a high level of concentration and a sense of control which could both stimulate performance (Barthelmäs & Keller, 2020; Engeser & Rheinberg, 2008). Regardless of the expectations, the findings of this research did not show evidence for a mediating effect of flow. The first ANOVA analysis revealed that voice usage changed significantly over time for both conditions. However, when the flow experience was included as a covariate in ANCOVA, the effect on voice usage was no longer significant. Therefore, flow may have accounted for some of the variance in voice usage scores. This implied that a mediation effect could be possible. However, the mediation analysis revealed no evidence for the mediating effect of flow. This also indicates that the observed changes in voice usage over time were not indirectly affected by flow. In addition to the mediation analysis, results of the ANOVA analyses revealed that even though the flow state of participants in the second presentation was significantly lower, filler words and appropriate use of inserted pauses scores were significantly higher than the delayed feedback condition. Contrary to the hypothesis, this reveals that even though participants experience lower flow states, voice usage skills can be positively affected during public speaking.

These results do not align with previous research by Kang et al. (2022) on the mediating effect of flow. However, the context of their study differed, and they investigated the mediating effect of flow between media characteristics as the independent variable and learning transfer as the dependent variable. Media characteristics did not include immediate or delayed feedback, but elements contributing to making the VR environment close to reality. Moreover, the researchers used a different instrument to capture the flow experience, including the instrument of Jackson and Marsh (1996) measuring the original nine components of flow. It could be the case that the instrument used in this current study might not have captured the whole experience of flow, resulting in, for instance, the flow experience being less or stronger during a presentation than revealed. Whether the instrument affected the results of the current study remains questionable, as studies (e.g., Delle Fave et al., 2011; Heutte et al., 2021; Norsworthy et al., 2023) suggest different methods. Researchers such as Delle Fave et al. (2011) state that measuring all nine components is crucial for capturing flow, while researchers such as Heutte et al. (2021) and Norsworthy et al. (2023) assume that measuring a few components is sufficient and easier to capture an overall flow experience. Concretely, no evidence was found in the current study for flow as a mediating factor, as was captured by Kang et al. (2022), possibly due to differences in contexts, variables, or measurement instruments. However, the role of the instrument remains uncertain, as several

researchers (e.g., Delle Fave et al., 2011; Heutte et al., 2021; Norsworthy et al., 2023) suggest different instruments for capturing flow effectively.

Even though the mediating effect is not captured in this study, it remains unclear why the immediate feedback condition, despite a weaker experience of flow, led to greater improvements in voice usage elements such as filler words and appropriate use of inserted pauses compared to the delayed feedback condition in the second presentation. A possible explanation is the uncertainty of whether the flow experience contributes to performance improvements. In addition, while Barthelmäs and Keller (2020) suggest that flow may influence performance, the researchers also acknowledge uncertainty about whether this relationship exists or is reversed, with performance influencing the flow experience. For instance, an important precondition of inducing the flow experience is a perceived fit between skills and demands, acknowledged by different research (e.g., Baumann et al., 2016; Keller & Bless, 2008; Keller & Blomann, 2008; Moneta & Csikszentmihalyi, 1996). Thus, if someone feels they have the skills to meet the challenges within the VR environment, they may perceive themselves as more competent, which potentially increases the likelihood of performing well on the task and improving flow. Schüler (2007) revealed that the flow experience influences performance, although the context differed, and their study considered measuring the "perceived balance between skills and challenges" into consideration. In contrast, the current study raises questions about the effect of performance on the flow experience: flow scores were weaker during the second presentation for the immediate feedback condition, yet performance improved. Interestingly, in the third presentation, without immediate feedback, flow scores increased, but voice usage scores remained relatively stable. It could be that lower flow scores, due to immediate feedback, negatively influenced the perceived balance between skills and demands for the third presentation, impacting performance. However, this was not measured. Ultimately, the results suggest that immediate feedback may positively affect performance more strongly than a lower state of flow on performance, though this conclusion lacks support from existing literature.

In sum, although existing research (Barthelmäs & Keller, 2020; Engeser & Rheinberg, 2008) revealed positive associations between flow and performance, the current study found no evidence that reduced flow experiences negatively influence voice usage or mediate the effect of immediate feedback on voice usage. One possible explanation is the inconsistent findings regarding the connection between flow and performance. Moreover, studies (e.g., Schüler, 2007) that revealed a positive relationship employed alternative instruments in different contexts.

Reliability and Validity of Instruments

Regarding the Psychological Flow Scale (PFS) questionnaire, a reliability and factor analysis was conducted to assess the reliability and validity of the instrument. There might have been a few

aspects that could affect the reliability and validity of this instrument. Particularly, the novelty of the instrument as it was recently developed in 2023, and translating the instrument into Dutch as the original PFS questionnaire is intended for English-speaking adults. Nevertheless, the items were highly correlated to the respective factor and Cronbach's alpha revealed that each factor was very reliable.

As mentioned in the method section, changes were made to the original voice use rubric to make the instrument more suitable for the current study. Specifically, the category "volume" was excluded from the rubric, and the category "pauses" of two criteria was merged into one category "appropriate use of inserted pauses". Additional explanations for scoring the categories were added as guidelines for the researcher and second assessor. Interrater reliability showed that the initial round of assessing three random presentations showed a .39 Cohen's Kappa statistic, indicating a fair/moderate agreement. Together with the second assessor, these scores have been discussed and adjusted. Hereafter, the adjusted scores were used for another interrater reliability test and showed a Kappa statistic of .90. A possible explanation for the big difference between the statistics may be due to the complexity of the assessment activity. For instance, assessing voice usage includes a definitive decision based on a period when fluctuations in each voice use element can appear. Moreover, Cho and Dewaele (2021) revealed that because of systematic differences of individuals, each listener could perceive the voice usage skills, such as intonation, differently. Adding an exhaustive description for each grade may have increased the complexity of the rubric. In addition, critically evaluating each voice usage skill, required the second assessor to have a thorough understanding of each skill, which was further enhanced through discussions of the results.

Theoretical Implications

This study expands on research outcomes by demonstrating the effect of immediate feedback on the flow experience within a VR public speaking training, which has not been investigated to the best of current knowledge. In contrast to research by Belboukhaddaoui and van Ginkel (2019), who revealed no difference between the effect of immediate versus delayed feedback on performance in VR public speaking training, the results of the current study revealed that delayed consistently enhances overall voice usage and flow experience each round, outperforming immediate feedback. Conversely, voice usage elements such as filler words and appropriate use of inserted pauses improved while receiving immediate feedback, revealing the positive effect that immediate feedback can have on performance, as shown earlier by King et al. (2000) in a study on the effect of feedback in real-life public speaking settings. However, this finding raises questions about whether – in line with the cognitive load theory – there might be a threshold for immediate feedback that can be processed to learn and improve voice usage skills during public speaking.

The current study also contributes to research on the effect of flow on performance. In addition to the results of research by Keller and Bless (2008) and Keller and Blomann (2008) who found no relationship between flow and performance, the current study revealed that although a lower flow experience occurs, performance can improve. It remains uncertain if the relationship between performance and flow is nonexistent, considering that after receiving delayed feedback the flow experience and voice usage improved. Lastly, the current study contributes by testing the applicability of the recently developed PFS questionnaire by Norsworthy et al. (2023) in another context, as recommended by the authors.

Practical Implications

From a practical standpoint, the outcomes of the current study offer valuable insights for educators and VR developers to enhance the learning experience of the individual. Even though there is not enough evidence that a lower flow experience influences the voice usage of participants, it is evident that receiving immediate feedback causes lower levels of absorption, effortless control, and intrinsic reward. Furthermore, according to the creator of the PFS questionnaire, Norsworthy et al. (2023) there is a negative correlation between these dimensions and the perceived stress and anxiety of the individual. Similarly, delayed feedback can improve the voice usage of the individual even more than immediate feedback. Therefore, one practical guideline is to critically question the effectiveness of immediate feedback and consider delayed feedback as a potentially more beneficial alternative within VR public speaking training. Moreover, public speaking can be a stressful or anxious activity (Tsang, 2020) and immediate feedback might contribute to this as well.

Alternatively, VR developers could consider revising the content included in the immediate feedback function of VR public speaking training. For instance, the distractive nature of immediate can be avoided if the presenter can find immediate feedback only when they need it. An example of the effectiveness of this intervention was shown in research by Tanveer et al. (2015). Moreover, as mentioned before, the presenter should have a clear understanding of the strengths and weaknesses of their performance after the presentation. Thus, developers could construct an immediate feedback system that reflect these insights.

Limitations

Several limitations need to be considered for interpreting the results of this study. Specifically, the VR software did not derive immediate nor delayed feedback on “articulation”, although this skill was assessed and is included on the voice use rubric. Therefore, fluctuations in the level of performance of this skill could be due to the novelty effect, as most participants indicated

that they were not exposed to Ovation before attending the research. The novelty effect suggests if a tool or activity is new to someone, it can positively affect behavior, especially in contexts of innovative technologies (Elston, 2021). Additionally, the novelty effect cannot be ruled out for other voice usage elements and the flow experience and can be considered a limitation of this study. Potential solutions include longer trials, involving participants familiar with the training tool (van Ginkel et al., 2019) or by including a third control condition in which participants are not exposed to immediate or delayed feedback.

A second limitation includes the dependency on technology of the intervention's (delayed or immediate feedback) quality, which could have affected the participants' performance. For instance, a participant with an "extremely" fast speaking rate based on the voice use rubric received an almost "perfect score" on speaking rate through delayed feedback. Based on the delayed feedback, the participant might have perceived the speaking rate as sufficient, preventing them from improving this skill. A similar finding was found by Cherner and Fegely (2023), who revealed that participants found generated feedback from a public speaking technology unreliable. Successful technological development comes with connecting non-research organizations to organizations that conduct research, such as universities (Hall et al., 2001). Software developers may not have investigated the conditions under which a voice usage skill can be considered a "perfect score". Therefore, developers must constantly update, improve, and evaluate feedback practices (Cherner & Fegely, 2023).

The final limitation refers to the language displayed in the VR software and processed by the VR software. To be concrete, the VR interface, including immediate feedback messages and the delayed feedback report, were displayed in English and could not be supported in Dutch. As most participants of the current study were Dutch-speaking, this could have affected participants' comprehension and responses to feedback as English most probably would not have been their native language. A possible solution is to provide detailed instructions to participants before the first presentation, defining the expected performance and evaluating their understanding (Mak, 2019), for instance, by explaining what (sufficient) "pitch variation" means. Another challenge regarding this limitation is that the technology was configured to provide feedback based on the English spoken language, this could have affected the output of immediate and delayed feedback provided to Dutch speakers.

Suggestions future research

During the data collection, the software used in this study introduced a beta functionality. This feature includes a generative artificial intelligence (GenAI) feature, which can provide more in-depth feedback after a presentation. Chen et al. (2024) revealed that GenAI can positively affect the

speaking performance of the individual. Future research could investigate whether this kind of feedback could be a solution for poor feedback strategies supporting the development of voice usage skills and if it can be an effective alternative for immediate feedback.

Another suggestion for future research is to investigate personal perceptions before the individual executes a task where flow could be experienced (e.g., public speaking). As previously mentioned, the perceived balance between skills and challenges could be a precondition to experience flow, as stated by Barthelmäs and Keller (2020). To investigate the perception before an activity, Csikszentmihalyi (1975) suggests capturing personality traits that can cause someone to overestimate or underestimate the objective challenges of the task. A method applied by Schüler (2007) involves asking participants to extend the sentence further "For me, personally, the current challenge is…" (p. 220), by choosing a number from 1 to 9. Specifically, nine indicate that the challenges are perceived as too high, and one indicates that the challenges perceived are too low for the individual. The result of capturing these perceptions could give insights into the reason why someone experiences low or higher flow states. In light of the current study, these insights could explain whether this precondition had a role in higher or lower flow states for the immediate feedback condition.

The final suggestion is to investigate the relationship between the flow experience, public speaking anxiety, and immediate feedback. As mentioned before, experiencing anxiety is a common challenge for individuals during public speaking, affecting their performance. Factors affecting public speaking anxiety have been investigated by several researchers (Lee et al., 2002; Tsang, 2020; Wang et al., 2020). Practicing public speaking in a safe environment, such as VR, could reduce public speaking anxiety, as revealed by Botella et al. (2000). However, as Norsworthy et al. (2021) stated, a negative relationship can exist between the flow dimensions and anxiety generally. This could potentially mean that immediate feedback might cause higher states of anxiety, which could influence performance. Investigating this aspect could expand existing research on public speaking anxiety, exploring the moderating factor of flow between immediate feedback and public speaking anxiety.

Conclusion

To conclude, the current study expands on research outcomes on the experience of flow and effectiveness of (formative) feedback timing, by investigating the effect of immediate versus delayed feedback on voice usage in VR public speaking training. Results revealed that immediate and delayed feedback both contribute to improving voice usage elements, whereas delayed feedback has a stronger impact. Moreover, immediate feedback negatively affects the flow experience, causing lower levels of absorption, effortless control, and intrinsic reward. Despite a weaker experience of

flow, immediate feedback led to greater improvements in filler words and appropriate use of inserted pauses compared to the delayed feedback condition in the second presentation. This insight reveals that while researchers (e.g., Schüler, 2007; Engeser & Rheinberg, 2008; Barthelmäs & Keller, 2020) state that the experience of flow can affect performance, the results of this study do not provide empirical support. Concretely, the current study offers relevant insights for educators and VR developers to enhance the learning experience of the speaker in VR. They can critically question the effectiveness of immediate feedback and might favor delayed feedback to improve voice usage skills. Moreover, strategies such as “sparse feedback” could be developed, potentially reducing the distractive nature of immediate feedback.

Reference List

Amende, R. (2024). *Effects of an instructional video on the quality of voice use in professional presentations* [Master's thesis, University of Twente]. https://purl.utwente.nl/essays/100649

Barhelmäs, M., Keller, J. (2020). *Antecedents, boundary conditions and consequences of flow.* Advances in Flow Research. (2), 71-107. https://doi.org/10.1007/978-3-030-53468-4_3

Baumann, N., Lürig, C., & Engeser, S. (2016). *Flow and enjoyment beyond skill-demand balance: The role of game pacing curves and personality.* Motivation and Emotion, *40*(4), 507–519. https://doi.org/10.1007/s11031-016-9549-7

Belboukhaddaoui, I., & van Ginkel, S. (2019). *Fostering oral presentation skills by the timing of feedback: an exploratory study in virtual reality.* Research on Education and Media, *11*(1), 25–31. https://doi.org/10.2478/rem-2019-0005

Blunck, P. M. (1997). *A Communication Competency Assessment Framework: A Literature Review of Communication Competency and Assessment.*

Botella, C., Baños, R., Guillén, V., Perpiña, C., Alcañiz, M., & Pons, A. (2000). *Telepsychology: Public speaking fear treatment on the internet.* CyberPsychology & Behavior, *3*(6), 959–968. https://doi.org/10.1089/109493100452228

Bower, M., Cavanagh, M., Moloney, R., & Dao, M. M. (2011). *Developing communication competence using an online video reflection system: Pre-service teachers' experiences.* Asia-Pacific Journal of Teacher Education, *39*(4), 311–326. https://doi.org/10.1080/1359866X.2011.614685

Chen, A., Jia, J., Li, Y., & Fu, L. (2024*). Investigating the effect of role-play activity with genAI agent on EFL Students' Speaking Performance*. Journal of Educational Computing Research, 1–27. https://doi.org/10.1177/07356331241299058

Cherner, T., & Fegely, A. (2023*). AI-powered presentation platforms for improving public speaking skills: Takeaways and suggestions for improvement.* Journal of Interactive Learning Research, *34*(2), 339–367. https://www.researchgate.net/publication/372775851

Cho, C. M., & Dewaele, J. M. (2021*). A crosslinguistic study of the perception of emotional intonation.* Studies in Second Language Acquisition, *43*(4), 870–895. https://doi.org/10.1017/S0272263120000674

Chollet, M., Marsella, S., & Scherer, S. (2022). *Training public speaking with virtual social interactions: effectiveness of real-time feedback and delayed feedback.* Journal on Multimodal User Interfaces, *16*(1), 17–29. https://doi.org/10.1007/s12193-021-00371-1

Chollet, M., Wörtwein, T., Morency, L. P., Shapiro, A., & Scherer, S. (2015). *Exploring feedback strategies to improve public speaking: An interactive virtual audience framework.* International

Joint Conference on Pervasive and Ubiquitous Computing, 1143–1154. https://doi.org/10.1145/2750858.2806060

Clark, H. H., & Fox Tree, J. E. (2002). *Using uh and um in spontaneous speaking.* Cognition*, 84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Corbett, A. T., & Anderson, J. R. (2001). *Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes.* Association for Computing Machinery, *3*(1), 245-252. https://doi.org/10.1145/365024.365111

Csikszentmihalyi, M. (1975). *Play and Intrinsic Rewards.* Journal of Humanistic Psychology, *15*(3), 41-63. https://doi-org.ezproxy2.utwente.nl/10.1177/002216787501500306

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*

Csikszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention.* HarperCollins Publishers, 107–126.

Damian, I., Tan, C. S. S., Baur, T., Schöning, J., Luyten, K., & André, E. (2015). *Augmenting social interactions: Realtime behavioural feedback using social signal processing techniques.* Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15), 565–574. Association for Computing Machinery. https://doi.org/10.1145/2702123.2702314

De Baer, S., & Feryn, T. (2020). Intonatie in het Nederlands: De invloed van leeftijd en geslacht. https://lib.ugent.be/catalog/rug01:002836368

De Grez, L., & Leuven, K. U. (2009). *Optimizing the instructional environment to learn presentation skills.* https://www.researchgate.net/publication/312493772

De Grez, L., Valcke, M., & Roozen, I. (2009a). *The impact of an innovative instructional intervention on the acquisition of oral presentation skills in higher education.* Computers and Education, *53*(1), 112–120. https://doi.org/10.1016/j.compedu.2009.01.005

De Grez, L., Valcke, M., & Roozen, I. (2009b). *The impact of goal orientation, self-reflection and personal characteristics on the acquisition of oral presentation skills.* European Journal of Psychology of Education, *24*(3), 293–306. https://doi.org/10.1007/BF03174762

Delle Fave, A., Massimini, F., & Bassi, M. (2011). *Psychological selection and optimal experience across cultures: Social empowerment through personal growth*. New York: Springer. https://doi.org/10.1007/978-90-481-9876-4

Delucenay, A. J., Conn, K. M., & Corigliano, A. (2017). *An evaluation of the impact of immediate compared to delayed feedback on the development of counselling skills in pharmacy students*. Pharmacy Education, *17*(1), 322–328. https://www.researchgate.net/publication/320617010_An_evaluation_of_the_impact_of_imm

ediate_compared_to_delayed_feedback_on_the_development_of_counselling_skills_in_pharmacy_students

Dihoff, R. E., Brosvic, G. M., Epstein, M. L., & Cook, M. J. (2004). *Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback*. The Psychological Record, *54*(2), 207–231. https://psycnet.apa.org/record/2004-13551-004

Dunbar, N. E., Brooks, C. F., & Kubicka-Miller, T. (2006). *Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills.* Innovative Higher Education, *31*(2), 115–128. https://doi.org/10.1007/s10755-006-9012-x

Elbert, S. P., & Dijkstra, A. (2014). *An experimental test of the relationship between voice intonation and persuasion in the domain of health.* Psychology and Health, *29*(9), 1014–1031. https://doi.org/10.1080/08870446.2014.903482

Elston, D. M. (2021). *The novelty effect.* Journal of the American Academy of Dermatology, *85*(3), 565–566. https://doi.org/10.1016/j.jaad.2021.06.846

Engeser, S., Peifer., C. (2021). *Theoretical integration and future lines of flow research.* Advances in Flow Research. (2), 417 – 439. https://doi.org/10.1007/978-3-030-53468-4_16

Engeser, S., & Rheinberg, F. (2008). *Flow, performance and moderators of challenge-skill balance.* Motivation and Emotion, *32*(3), 158–172. https://doi.org/10.1007/s11031-008-9102-4

Engeser, S., Schiepe-Tiska, A., Peifer, C. (2012). *Historical lines and an overview of current research on flow.* Advances in Flow Research. (2), 1-29. https://doi.org/10.1007/978-3-030-53468-4_1

Gaggioli, A., Bassi, M., & Delle Fave, A. (2003). *Quality of experience in virtual environments*. Being there: Concepts, effects and measurement of user presence in synthetic environments, 121–135. https://www.researchgate.net/publication/255171801

Guyer, J., & Fabrigar, L. (2017). *Vocal confidence and persuasion: Speech rate affects amount of processing under moderate elaboration*. https://www.researchgate.net/publication/312314290

Guyer, J. J., Briñol, P., Vaughan-Johnston, T. I., Fabrigar, L. R., Moreno, L., & Petty, R. E. (2021). *Paralinguistic features communicated through voice can affect appraisals of confidence and evaluative judgments.* Journal of Nonverbal Behavior, *45*(4), 479–504. https://doi.org/10.1007/s10919-021-00374-2

Hall, A., Bockett, G., Taylor, S., Sivamohan, M. V. K., & Clark, N. (2001). *Why research partnerships really matter: Innovation theory, institutional arrangements and implications for developing new technology for the poor*. World Development, *29*(5), 783–797. https://doi.org/10.1016/S0305-750X(01)00004-3

Heutte, J., Fenouillet, F., Martin-Krumm, C., Gute, G., Raes, A., Gute, D., Bachelet, R., & Csikszentmihalyi, M. (2021). *Optimal experience in adult learning: conception and validation of*

the flow in education scale (EduFlow-2). Frontiers in Psychology, *12*, 1–12. https://doi.org/10.3389/fpsyg.2021.828027

House, J. (2006). *Constructing a context with intonation*. Journal of Pragmatics, *38*(10), 1542–1558. https://doi.org/10.1016/j.pragma.2005.07.005

Jackson, S. A., & Marsh, H. W. (1996). *Development and validation of a scale to measure optimal experience: The flow state scale*. Journal of Sport and Exercise Psychology, *18*(1), 17–35. https://doi.org/10.1123/jsep.18.1.17

Jiang, X., & Pell, M. D. (2017). *The sound of confidence and doubt.* Speech Communication, *88*, 106–126. https://doi.org/10.1016/j.specom.2017.01.011

Kang, Y. G., Song, H. D., Yun, H., & Jo, Y. (2022). *The effect of virtual reality media characteristics on flow and learning transfer in job training: The moderating effect of presence.* Journal of Computer Assisted Learning, *38*(6), 1674–1685. https://doi.org/10.1111/jcal.12702

Keller, J., & Bless, H. (2008). *Flow and regulatory compatibility: An experimental approach to the flow model of intrinsic motivation.* Personality and Social Psychology Bulletin, *34*(2), 196–209. https://doi.org/10.1177/0146167207310026

Keller, J., & Blomann, F. (2008). *Locus of control and the flow experience: An experimental analysis.* European Journal of Personality, *22*(7), 589–607. https://doi.org/10.1002/per.692

Kerby, D., & Romine, J. (2009). *Develop oral presentation skills through accounting curriculum design and course-embedded assessment.* Journal of Education for Business, *85*(3), 172–179. https://doi.org/10.1080/08832320903252389

Kimble, C. E., & Seidel, S. D. (1991). *Vocal signs of confidence.* Journal of Nonverbal Behavior, 15(2), 99–105. https://doi.org/10.1007/BF00998265

King, P. E., Young, M. J., & Behnke, R. R. (2000). *Public speaking performance improvement as a function of information processing in immediate and delayed feedback interventions.* Communication Education, *49*(4), 365–374. https://doi.org/10.1080/03634520009379224

Knowlton, G. E., & Larkin, K. T. (2006). The influence of voice volume, pitch, and speech rate on progressive relaxation training: Application of methods from speech pathology and audiology. Applied Psychophysiology Biofeedback, *31*(2), 173–185. https://doi.org/10.1007/s10484-006-9014-6

Kothgassner, O. D., Felnhofer, A., Beutl, L., Hlavacs, H., Lehenbauer, M., & Stetina, B. (2012). *A virtual training tool for giving talks.* Entertainment Computing – ICEC 2012, 53–66. https://doi.org/10.1007/978-3-642-33542-6_5

Kulik, J. A., & Kulik, C. L. C. (1988). *Timing of feedback and verbal learning.* Review of Educational Research, *58*(1), 79–97. https://doi.org/10.2307/1170349

Laske, M. M., & DiGennaro Reed, F. D. (2024). *Um, so, like, do speech disfluencies matter? A parametric evaluation of filler sounds and words.* Journal of Applied Behavior Analysis, *57*(3), 574–583. https://doi.org/10.1002/jaba.1093

Lee, J. M., Ku, J. H., Jang, D. P., Kim, D. H., Choi, Y. H., Kim, I. Y., & Kim, S. I. (2002). *Virtual reality system for treatment of the fear of public speaking using image-based rendering and moving pictures.* Cyberpsychology & Behavior. *5*(3), 191–195. https://doi.org/10.1089/109493102760147169

Levelt, W. J. M. (1989). *Speaking: from intention to articulation*. 1–584. https://doi.org/10.7551/mitpress/6393.001.0001

Mak, S. H. (2019). *Analysing the needs of EFL/ESL learners in developing academic presentation competence.* RELC Journal, 52(3), 379–396. https://doi.org/10.1177/0033688219879514

Mcalpine, L. (2004). *Designing learning as well as teaching: A research-based model for instruction that emphasizes learner practice*. Active Learning in Higher Education, *5*(2), 119–134. https://doi.org/10.1177/1469787404043809

Metcalfe, J., Kornell, N., & Finn, B. (2009). *Delayed versus immediate feedback in children's and adults' vocabulary learning.* Memory and Cognition, *37*(8), 1077–1087. https://doi.org/10.3758/MC.37.8.1077

Miller. G. A. (1956). *The magical number seven, plus or minus two: Some limits on our capacity for processing information.* The Psychological Review. *63*(2), 81-97. https://doi.org/10.1037/h0043158

Moneta, G. B., & Csikszentmihalyi, M. (1996). *The Effect of Perceived Challenges and Skills on the Quality of Subjective Experience.* Journal of Personality, *64*(2), 275–310. https://doi.org/10.1111/j.1467-6494.1996.tb00512.

Moneta., G. B. (2021). *On the conceptualization and measurement of flow.* (2), 31-69. https://doi.org/10.1007/978-3-030-53468-4_2

Montes, C. C., Heinicke, M. R., & Geierman, D. M. (2019). *Awareness training reduces college students' speech disfluencies in public speaking.* Journal of Applied Behavior Analysis, *52*(3), 746–755. https://doi.org/10.1002/jaba.569

Moore, D. L., Hausknecht, D., & Thamodaran, K. (1986). *Time compression, response opportunity, and persuasion.* Journal of consumer Research, *13*(1), 85–99. https://www.jstor.org/stable/2489288

Morreale, S., Hackman, M., Ellis, K., King, K., Meade, P. A., & Pinello-Tegtmeier, L. (1993). *Assessing Communication Competency in the Interpersonal Communication Course: A Laboratory-Supported Approach*. 1–20. https://eric.ed.gov/?id=ED366024

Narciss, S. (2008). *Feedback Strategies for Interactive Learning Tasks.* 125–143. https://www.researchgate.net/publication/233833269

Nicol, D. J., & Macfarlane-Dick, D. (2006). *Formative assessment and self-regulated learning: A model and seven principles of good feedback practice.* Studies in Higher Education*, 31*(2), 199-218. https://doi.org/10.1080/03075070600572090

Nishimura, Y., & Hashida, T. (2018). *Recognition and feedback of vowel utterance with a good mouth shape based on sensing platysma muscle bulging.* ACM International Conference Proceeding Series. https://doi.org/10.1145/3174910.3174944

Norsworthy, C., Dimmock, J. A., Miller, D. J., Krause, A., & Jackson, B. (2023). *Psychological flow scale (PFS): Development and preliminary validation of a new flow instrument that measures the core experience of flow to reflect recent conceptual advancements.* International Journal of Applied Positive Psychology, *8*(2), 309–337. https://doi.org/10.1007/s41042-023-00092-8

Norsworthy, C., Jackson, B., & Dimmock, J. A. (2021). *Supplemental material for advancing our understanding of psychological flow: A scoping review of conceptualizations, measurements, and applications.* Psychological Bulletin, *147*(8), 806–827. https://doi.org/10.1037/bul0000337.supp

Okrasa, M., Wozniak, M. P., Walczak, A., Bartłomiejczyk, N., Müller, H., Grudzień, K., & Romanowski, A. (2022). *Supporting self-development of speech delivery for education professionals.* ACM International Conference Proceeding Series, 251–253. https://doi.org/10.1145/3568444.3570588

Ortiz, L. A., Region-Sebest, M., & MacDermott, C. (2016). *Employer perceptions of oral communication competencies most valued in new hires as a factor in company success.* Business and Professional Communication Quarterly, *79*(3), 317–330. https://doi.org/10.1177/2329490615624108

Pearce, J. M., Ainley, M., & Howard, S. (2005). *The ebb and flow of online learning.* Computers in Human Behavior, *21*(5), 745–771. https://doi.org/10.1016/j.chb.2004.02.019

Phye, G. D., & Andre, T. (1989). *Delayed retention effect: Attention, perseveration, or both?* Contemporary Educational Psychology, *14*, 173–185. https://doi.org/10.1016/0361-476X(89)90035-0

Poeschl, S. (2017). *Virtual reality training for public speaking-A QUEST-VR framework validation. Frontiers in ICT*, *4*(13), 1–13. https://doi.org/10.3389/fict.2017.00013

Rhodes, S. C., & Frandsen, K. D. (1975). *Some effects of instruction in feedback utilization on the fluency of college students' speech.* Speech Monographs, 42, 83–89. https://doi.org/10.1080/03637757509375881

Scherer, K. R., & Wolf, J. J. (1973). *The voice of confidence: Paralinguistic cues and audience evaluation.* Journal of Research in Personality, *7*(1), 31–44. https://doi.org/10.1016/0092-6566(73)90030-5

Schnotz, W., & Kürschner, C. (2007). *A reconsideration of cognitive load theory.* Educational Psychology Review, *19*(4), 469–508. https://doi.org/10.1007/s10648-007-9053-4

Schroth, M. L., & Lund, E. (1993). *Role of delay of feedback on subsequent pattern recognition transfer task.* Contemporary Educational Psychology, *18*, 15–22. https://doi.org/10.1006/ceps.1993.1003

Schüler, J. (2007). *Arousal of flow experience in a learning setting and its effects on exam performance and affect.* Zeitschrift Fur Padagogische Psychologie, *21*(3–4), 217–227. https://doi.org/10.1024/1010-0652.21.3.217

Shute, V. J. (2008). *Focus on formative feedback.* Review of Educational Research, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Si Na, K., Vei Siang, C., Mohammed, F., Tasir, Z., Isham, M. I. M., & Abas, M. A. (2020). *Virtual reality application integrated with learning analytics for enhancing English pronunciation: A conceptual framework.* 2020 IEEE Conference on e-Learning, e-Management and e-Services (IC3e). https://doi.org/10.1109/IC3e50159.2020.9288478

Siddiqui, H., Irfan, H., Lakhani, A. S., Ahmed, B., Shaikh, S., Movania, M. M., & Farhan, M. (2023). *Manifest: Public speaking training using virtual reality.* Proceedings - 2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct, ISMAR-Adjunct 2023, 468–473. https://doi.org/10.1109/ISMAR-Adjunct60411.2023.00102

Smith, C. M., & Sodano, T. M. (2011). *Integrating lecture capture as a teaching strategy to improve student presentation skills through self-assessment.* Active Learning in Higher Education, *12*(3), 151–162. https://doi.org/10.1177/1469787411415082

Sokoreva, T., & Shevchenko, T. (2022). *The role of pause in interaction: A case of polylogue*. Springer International Publishing. https://doi.org/10.1007/978-3-031-20980-2

Swann, C., Piggott, D., Schweickle, M., & Vella, S. A. (2018). *A review of scientific progress in flow in sport and exercise: Normal science, crisis, and a progressive shift.* Journal of Applied Sport Psychology, *30*(3), 249–271. https://doi.org/10.1080/10413200.2018.1443525

Sweller, J., & Chandler, P. (1991). *Evidence for cognitive load theory*. Cognition and Instruction, *8*(4), 351–362. https://doi.org/10.1207/s1532690xci0804_5

Sweller, J., & Chandler, P. (1994). *Why some material is difficult to learn*. Cognition and Instruction, *12*(3), 185-233. https://doi.org/10.1207/s1532690xci1203_1

Tanveer, M. I., Lin, E., & Hoque, M. (2015). *Rhema: A real-time in-situ intelligent interface to help people with public speaking.* International Conference on Intelligent User Interfaces, Proceedings IUI, 2015-January, 286–295. https://doi.org/10.1145/2678025.2701386

Triberti, S., Di Natale, A. F., Gaggioli, A. (2021). *Flowing technologies: The role of flow and related constructs in human-computer interaction.* Advances in Flow Research. *2*, 393 – 416. https://doi.org/10.1007/978-3-030-53468-4_15

Tsang, A. (2017). *Enhancing learners' awareness of oral presentation (delivery) skills in the context of self-regulated learning.* Active Learning in Higher Education, *21*(1), 39–50. https://doi.org/10.1177/1469787417731214

Tsang, A. (2018). *Positive effects of a program on oral presentation skills: high- and low-proficient learners' self-evaluations and perspectives.* Assessment and Evaluation in Higher Education, *43*(5), 760–771. https://doi.org/10.1080/02602938.2017.1407917

Tsang, A. (2020). *The relationship between tertiary-level students' self-perceived presentation delivery and public speaking anxiety: A mixed-methods study.* Assessment and Evaluation in Higher Education, *45*(7), 1060–1072. https://doi.org/10.1080/02602938.2020.1718601

van Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2015). *Towards a set of design principles for developing oral presentation competence: A synthesis of research in higher education.* Educational Research Review. *14*, 62–80. https://doi.org/10.1016/j.edurev.2015.02.002

van Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2017). *The impact of the feedback source on developing oral presentation competence.* Studies in Higher Education, *42*(9), 1671–1685. https://doi.org/10.1080/03075079.2015.1117064

van Ginkel, S., Gulikers, J., Biemans, H., Noroozi, O., Roozen, M., Bos, T., van Tilborg, R., van Halteren, M., & Mulder, M. (2019). *Fostering oral presentation competence through a virtual reality-based task for delivering feedback.* Computers and Education, *134*, 78–97. https://doi.org/10.1016/j.compedu.2019.02.006

Van Merriënboer, J. J. G., & Sweller, J. (2010). *Cognitive load theory in health professional education: Design principles and strategies.* Medical Education, *44*(1), 85–93. https://doi.org/10.1111/j.1365-2923.2009.03498.x

Van Zant, A., & Berger, J. (2020). *How the voice persuades.* Journal of Personality and Social Psychology, *118*(4), 661-682. https://doi.org/10.1037/pspi0000193.supp

Volante, W. G., Cruit, J., Tice, J., Shugars, W., & Hancock, P. A. (2018). *Time flies: Investigating duration judgments in virtual reality.* Proceedings of the Human Factors and Ergonomics Society, *3*, 1777–1781. https://doi.org/10.1177/1541931218621403

Wang, J., Yang, H., Shao, R., Abdullah, S., & Sundar, S. S. (2020). *Alexa as coach: leveraging smart speakers to build social agents that reduce public speaking anxiety.* Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3313831.3376561

Weger, U. W., Meier, B. P., Robinson, M. D., Inhoff, A. W., & Weger, U. W. (2007). *Things are sounding up: Affective influences on auditory tone perception.* Psychon Bull Rev, *14*(3). https://doi.org/10.3758/BF03194100

Wennerstrom, A., & Siegel, A.F. (2003). *Keeping the floor in multiparty conversations: intonation, syntax, and pause.* Discourse Processes, *36(*2), 77–107.

https://doi.org/10.1207/S15326950DP3602_1

Appendices

Appendix A: Adjusted Voice Use Rubric

Voice Use Rubric

Presentation skills consist of several elements, including, for example, posture, eye contact, and voice use. A rubric has been developed for the latter, focusing on the categories of intonation, articulation, speaking speed, inserted pauses, and use of filler words. Each category is rated with grades 1 to 10, where 1 is absent and 10 is excellent.

To provide comprehensive guidelines for scoring the categories, each grade has been explained. It is important to note that these explanations serve as a framework and may not encompass all the criteria within each category. During the evaluation of presentations, these explanations are intended as general guidelines rather than exhaustive descriptions.

Category Intonation

Here the variation of the pitch of the voice while speaking is examined. The use of intonation allows the audience to organize information. A question or a statement can be represented by intonation. When asking a question, the voice often rises (upward inflexion). In addition, the speaker uses intonation to convey his emotions and attitude towards the information provided. The speaker will lower his voice when he wants to emphasize specific information to the audience (downward inflexion). The more enthusiastic the speaker is, the greater the intonation. When assessing intonation, a distinction is made between the variation of intonation and the appropriate use of intonation in the sentence or context.

Variation of Intonation

Listen for: changes in pitch throughout the presentation and enthusiasm of the speaker.

1 **Absent** No variation in pitch: the speaker speaks monotonously.	**2** **Failure** The variation in pitch is minor, rare, and barely noticeable, with a near-monotone delivery.	**3** **Poor** There are very few variations in pitch, with a near-monotone delivery.	**4** **Insufficient** Occasional variation in pitch: the speaker attempts to vary tone, but the variation is infrequent and minor.	**5** **Mediocre** Moderate variation in pitch: the speaker varies tone inconsistently and speaks slightly more monotonously than engaging.
6 **Sufficient** General variation in pitch: The speaker varies tone sufficiently maintaining a reasonable level of interest, occasionally speaking monotone.	**7** **Satisfactory** Good variation in pitch: The speaker's tone variation contributes to the presentation engagement, with infrequent moments of speaking monotonously.	**8** **Good** Good variation in pitch: the speaker effectively uses tone changes, with very few moments of speaking monotonously.	**9** **Very good** Very good variation in pitch throughout the presentation the speaker's tone changes are highly engaging.	**10** **Excellent** There is excellent variation in pitch: the speaker consistently employs dynamic and engaging pitch changes, with no areas needing improvement.

Appropriate use of intonation

Listen for: whether the intonation fits the sentence or context.

For example, the appropriate use of intonation is rated excellent (10) if the pitch goes up at the end of the sentence if the speaker asks a question during the presentation, and when the pitch goes down at the end of the sentence when the speaker wants to emphasize specific information.

1 **Absent** Incorrect intonation throughout the presentation.	**2** **Failure** Mostly incorrect intonation, correct use of intonation is minor, rare, and barely noticeable. Speaker uses upward and downward inflexion wrong.	**3** **Poor** Frequently incorrect intonation: very few correct uses of upward and downward inflexion.	**4** **Insufficient** Occasional incorrect uses of intonation: the correct use of upward and downward inflexion is infrequent and minor.	**5** **Mediocre** Moderate correct uses of intonation: the application of the correct use of intonation is inconsistent and slightly more incorrect.
6 **Sufficient** Generally correct intonation with occasional mistakes in	**7** **Satisfactory** Mostly correct intonation: any errors	**8** **Good** Correct intonation throughout the	**9** **Very good** Intonation is almost always correct: effectively enhances	**10** **Excellent** Intonation is consistently correct and highly effective: always matching the

using upward and downward inflexion.	are infrequent and minor.	presentation: very few mistakes.	the clarity and emphasis of the content. Mistakes are very rare and minor.	sentence structure and context, with no areas needing improvement.

Category Articulation

Articulation is the clear pronunciation of words in a sentence. Each syllable of a word is pronounced clearly. As a result, the speaker can be understood clearly and does not mumble.

Clear Pronunciation of Words

*Listen for: clear pronunciation of words***.**

1 **Absent** Words are unclear and mumbled: speaker is unintelligible.	**2** **Failure** Pronunciation is very unclear: most words are difficult to understand due to persistent mumbling or misarticulation.	**3** **Poor** Many words are unclear, very few can be understood frequent mumbling or misarticulation makes it often hard to understand the speaker.	**4** **Insufficient** Some unclear words, with only occasional instances where words can be understood, but these are infrequent and minor. Sometimes it is hard to understand the speaker.	**5** **Mediocre** Generally clear pronunciation, some instances of mumbling. The speech is somewhat understandable.
6 **Sufficient** Pronunciation is mostly clear: most words can be understood, minor mumbling sometimes affects comprehensibility.	**7** **Satisfactory** Clear pronunciation with infrequent and minor unclear words.	**8** **Good** The pronunciation is clear and mostly precise, making the speech easy to understand. Very few mistakes occur.	**9** **Very good** The pronunciation is very clear and precise, mistakes are very rare and minor: the speech is easy to understand.	**10** **Excellent** The pronunciation is consistently exceptionally clear and precise, making the speech very easy to understand, with no areas needing improvement.

Category Speed

When assessing speed, a distinction is made between the variation of speed and the appropriate use of speed in the presentation.

Variation in Speed

Listen for: the speaker varies in speed: then fast then slow again.

1 **Absent** No variation in speed: the speaker maintains a constant fast or slow pace throughout the presentation.	**2** **Failure** The variation in speed is minimal and barely noticeable, with a nearly uniform pace throughout.	**3** **Poor** The variation in speed is limited: the speaker changes pace a few times throughout the presentation.	**4** **Insufficient** The speaker demonstrates some variation in speed: however, the pace remains mostly consistent with only slight, infrequent changes.	**5** **Mediocre** There is moderate variation in speed: the speaker changes pace inconsistently, maintaining the same pace slightly more often than changing it.
6 **Sufficient** The speaker exhibits adequate variation in speed: there are noticeable changes in pace, though they may not be consistently executed.	**7** **Satisfactory** The speaker demonstrates variation in speed: changes in pace are more frequent.	**8** **Good** The variation in speed is good: the speaker effectively utilizes changes in pace.	**9** **Very good** Very good variation in speed: the speaker's changes in pace are very noticeable and dynamic.	**10** **Excellent** The variation in speed is excellent: the speaker consistently employs dynamic and engaging changes in pace throughout the presentation, with no areas needing improvement.

Appropriate Use of Speed

Appropriate use of speed considers whether the speed fits the context.

Listen for: If the speaker aims to provide the information, he speaks faster and if the speaker intends to make the audience think he speaks more slowly.

1 **Absent** The speaking speed is incorrect throughout the presentation, failing to align with the content and context.	**2** **Failure** The speaking speed is predominantly incorrect, with only rare instances of correct use, often leading to rushed delivery.	**3** **Poor** The speaking speed is frequently incorrect, with very few correct uses.	**4** **Insufficient** The speaker exhibits several incorrect uses of speed, with occasional instances of correct use.	**5** **Mediocre** There are slightly more incorrect applications of speed use, correct use is inconsistent.
6 **Sufficient** The speaking speed application is slightly more correct, with occasional errors.	**7** **Satisfactory** The speaking speed is mostly correct: any errors are infrequent and minor.	**8** **Good** The speaking speed is correct throughout the presentation, with very few mistakes.	**9** **Very good** The speaking speed is almost always correct throughout the presentation. Mistakes are very rare.	**10** **Excellent** The speaking speed is consistently correct and highly effective, aligning with the content and context, with no areas needing improvement.

Category pauses

The "pauses" category considers whether the speaker inserts pauses during his presentation. Pauses can, for example, be inserted to emphasize certain words or phrases to allow the audience to absorb the information provided.

Appropriate Use of Inserted Pauses

It is assessed whether the inserted pauses are applied appropriately. For example, inserted pauses are rated excellent (10) when the speaker takes small “breathing pauses” after completing a sentence, when the speaker takes a longer pause after changing to a topic, and when the speaker emphasizes certain words or phrases to give the audience a chance to absorb the information. Pauses can also be inserted because the speaker is searching for words, for example, using a filler word such as ‘ah’ or ‘uhm’. Such a pause can interrupt the flow of a presentation and this pause is then not considered appropriate.

Listen for: appropriate and strategic use of pauses (e.g. speaker is not searching for words or is using filler words when inserting a pause).

1 **Absent** The speaker does not pause appropriately. (e.g. does not pause after a sentence or when changing a topic. Pauses are due to using filler words)	**2** **Failure** Pauses are minor, rare, and barely noticeable, making the presentation hard to follow. The speaker inserts pauses mainly to search for words or use filler words.	**3** **Poor** The speaker pauses a very few times throughout the presentation. Pauses can often result from searching for words and using filler words.	**4** **Insufficient** The speaker exhibits several incorrect uses of pauses, with occasional instances of correct use. Pauses are often due to searching for words and using filler words. Parts of the presentation are hard to follow.	**5** **Mediocre** There are some correct uses of pauses, the application is inconsistent. Still, most pauses can result from searching for words and using filler words. Some parts of the presentation are still hard to follow.
6 **Sufficient** The speaker exhibits sufficient use of pauses but with some room for improvement in timing and frequency. Occasionally, pauses result from searching for words, and filler words.	**7** **Satisfactory** The speaker demonstrates good use of pauses after changing a sentence or topic. There are still minor errors where additional pauses should be inserted, or where filler words should be avoided.	**8** **Good** The use of pauses is good after changing a sentence or topic, enhancing the clarity of the presentation, with only very few areas for improvement.	**9** **Very good** The use of pauses is very good, consistently used throughout the presentation, with almost perfect timing and frequency. Areas for improvement are rare and minor.	**10** **Excellent** The use of pauses is excellent, the speaker always takes small breathing pauses after completing a sentence and longer pauses after changing a topic. No areas need improvement.

Category Filler Words

The speaker uses words, which fill the silence during a presentation. These words interrupt the flow of the presentation. With frequent use of filler words, it becomes difficult for the audience to follow the speaker, which makes it uncomfortable for the audience as well. For example, the speaker uses sounds like 'uhm', 'ah', click sounds, doubles words or uses meaningless words, e.g., 'like', 'I mean'. In this category, the fewer filler words used, the better the score.

Frequency of Filler Words

Listen for: the frequency of filler words (e.g. "uh", "uhm", "yes", "and", "so")

1 **Absent** The speaker uses more than 17 filler words per minute.	**2** **Failure** The speaker uses 15-16 filler words per minute.	**3** **Poor** The speaker uses 13-14 filler words per minute.	**4** **Insufficient** The speaker uses 11-12 filler words per minute.	**5** **Mediocre** The speaker uses 9-10 filler words per minute.
6 **Sufficient** The speaker uses 7-8 filler words per minute.	**7** **Satisfactory** The speaker uses 5-6 filler words per minute.	**8** **Good** The speaker uses 3-4 filler words per minute.	**9** **Very good** The speaker uses 1-2 filler words per minute.	**10** **Excellent** The speaker uses no filler words.

Appendix B: Voice Use Rubric

Presentation skills consist of several elements, including, for example, posture, eye contact and voice use. A rubric has been developed for the latter, focusing on the categories of intonation, articulation, speaking speed, volume, inserted pauses, and use of filler words. Each category is rated with grades 1 to 10, where 1 is absent and 10 is excellent.

Category

Intonation:

Here the variation of the pitch of the voice while speaking is examined. The use of intonation allows the audience to organise information. A question or a statement can be represented by intonation. When asking a question, the voice often rises (upward inflexion). In addition, the speaker uses intonation to convey his emotions and attitude towards the information provided. The speaker will lower his voice when he wants to emphasise specific information to the audience (downward inflexion). The more enthusiastic the speaker is, the greater the intonation. When assessing intonation, a distinction is made between the variation of intonation and the appropriate use of intonation in the sentence or context.

Variation of intonation is rated excellent (10) if the speaker varies the pitch throughout his presentation. The speaker is given an absent (1) if no variation in pitch can be heard. The speaker in this case speaks monotonously.

Appropriate use of intonation considers whether the intonation fits the sentence or context. For example, if the speaker asks a question, the intonation goes up at the end of the sentence. If the intonation is applied incorrectly during the presentation, then the use is rated absent (1), if the intonation is applied correctly throughout the presentation, then this category is rated excellent (10).

Variation of intonation:

1 Absent	2 Failure	3 Poor	4 Insufficient	5 Mediocre	6 Sufficient	7 Satisfactory	8 Good	9 Very good	10 Excellent

Appropriate use of intonation

1 Absent	2 Failure	3 Poor	4 Insufficient	5 Mediocre	6 Sufficient	7 Satisfactory	8 Good	9 Very good	10 Excellent

Articulation

Articulation is the clear pronunciation of words in a sentence. Each syllable of a word is pronounced clearly. As a result, the speaker can be understood clearly and does not mumble.

1 Absent	2 Failure	3 Poor	4 Insufficient	5 Mediocre	6 Sufficient	7 Satisfactory	8 Good	9 Very good	10 Excellent

Speed:

When assessing speed, a distinction is made between the variation of speed and the appropriate use of speed in the presentation.

During the presentation, it is noticeable that the speaker varies in speed; then fast then slow again. Appropriate use of speed considers whether the speed fits the context. If the speaker aims to provide the information, he speaks faster and if the speaker intends to make the audience think he speaks more slowly.

Variation of the speed

1 Absent	2 Failure	3 Poor	4 Insufficient	5 Mediocre	6 Sufficient	7 Satisfactory	8 Good	9 Very good	10 Excellent

Appropriate use of the speed

1 Absent	2 Failure	3 Poor	4 Insufficient	5 Mediocre	6 Sufficient	7 Satisfactory	8 Good	9 Very good	10 Excellent

Volume:

Volume is assessed as whether it is used appropriately. For example, if the speaker talks about whispering, it is appropriate to lower the voice volume.

Appropriate use of the volume

1 Absent	2 Failure	3 Poor	4 Insufficient	5 Mediocre	6 Sufficient	7 Satisfactory	8 Good	9 Very good	10 Excellent

Pauses:

The "pauses" category considers whether the speaker inserts pauses during his presentation. Pauses can, for example, be inserted to emphasise certain words or phrases to give the audience a chance to absorb the information provided.

It is further assessed whether these inserted pauses are applied appropriately. Pauses can also be inserted because the speaker is searching for words, for example. Such a pause can interrupt the flow of a presentation and this pause is then not considered appropriate.

Inserted pauses

1 Absent	2 Failure	3 Poor	4 Insufficient	5 Mediocre	6 Sufficient	7 Satisfactory	8 Good	9 Very good	10 Excellent

Appropriate use of the pauses

1 Absent	2 Failure	3 Poor	4 Insufficient	5 Mediocre	6 Sufficient	7 Satisfactory	8 Good	9 Very good	10 Excellent

Filler words:

The speaker uses words, which fill the silence during a presentation. These words interrupt the flow of the presentation. With frequent use of filler words, it becomes difficult for the audience to follow the speaker, which makes it uncomfortable for the audience as well. For example, the speaker uses sounds like 'uhm', 'ah', click sounds, doubles words or uses meaningless words, e.g., 'like', 'I mean'. In this category, the fewer filler words used, the better the score.

Use of filler words

1	2	3	4	5	6	7	8	9	10
Absent	Failure	Poor	Insufficient	Mediocre	Sufficient	Satisfactory	Good	Very good	Excellent

Appendix C: Flow Questionnaire

Psychological Flow Scale (PFS)

The below questions relate to the thoughts and feelings you may have experienced while taking part in your recent activity. There are no right or wrong answers. Think about how you felt during the event/activity, then answer the questions using the rating scale.

Please rate the questions below in relation to the most intense optimal moment you experienced in your given event.

		Strongly disagree	Neutral					Strongly agree
1	I was absorbed in the act/task	1	2	3	4	5	6	7
2	I was highly focused on the task/activity	1	2	3	4	5	6	7
3	All my attention was on the task/activity	1	2	3	4	5	6	7
4	I felt like I could easily control what I was doing	1	2	3	4	5	6	7
5	My actions flowed effortlessly	1	2	3	4	5	6	7

		Strongly disagree	Neutral					Strongly agree
6	There was a sense of fluidity to my actions	1	2	3	4	5	6	7
7	I found the experience rewarding	1	2	3	4	5	6	7
8	The experience felt satisfying	1	2	3	4	5	6	7
9	I would like the feeling of that experience again	1	2	3	4	5	6	7

Researcher notes:

Items 1-9 assess the flow experience. Average the nine scores to obtain a global flow score. Items 1-3 assess the dimension 'absorption'. Items 4-6 assess the dimension 'effort-less control'. Items 7-9 assess the dimension 'intrinsic reward'. Where possible it is advised to examine dimensional scores in addition to the global flow score. Avg means scores (for both global and the three dimensions) to be used for reporting.

Dutch-translated Questionnaire

De onderstaande vragen hebben betrekking op de gedachten en gevoelens die je mogelijk hebt ervaren tijdens het geven van je presentatie in Virtual Reality. Er zijn geen goede of foute antwoorden. Denk na over hoe je je voelde tijdens het geven van je presentatie en beantwoord vervolgens de vragen met behulp van de beoordelingsschaal. Beoordeel de onderstaande vragen in relatie tot het meest intense en optimale moment dat je hebt ervaren tijdens het geven van je presentatie.

	Sterk mee oneens			Neutraal			Sterk mee eens
	1	2	3	4	5	6	7
Ik ging volledig op in het geven van mijn presentatie.	○	○	○	○	○	○	○
Ik was zeer gefocust op het geven van mijn presentatie.	○	○	○	○	○	○	○
Al mijn aandacht was bij het geven van mijn presentatie.	○	○	○	○	○	○	○
Ik voelde dat ik makkelijk controle had over wat ik deed.	○	○	○	○	○	○	○
Mijn handelingen verliepen moeiteloos.	○	○	○	○	○	○	○
Ik had het gevoel dat wat ik deed heel makkelijk op elkaar volgde.	○	○	○	○	○	○	○
Ik vond de ervaring belonend.	○	○	○	○	○	○	○
Ik vond de ervaring voldoening geven.	○	○	○	○	○	○	○
Ik zou het gevoel van de ervaring opnieuw willen ervaren.	○	○	○	○	○	○	○