



Theses and Dissertations

2022-12-14

Examining the Willingness to Communicate (WTC) Scale with Advanced Foreign Language Learners

Colin Lilya
Brigham Young University

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Arts and Humanities Commons](#)

BYU ScholarsArchive Citation

Lilya, Colin, "Examining the Willingness to Communicate (WTC) Scale with Advanced Foreign Language Learners" (2022). *Theses and Dissertations*. 9783.

<https://scholarsarchive.byu.edu/etd/9783>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

Examining the WTC Scale with Advanced Foreign Language Learners

Colin Lilya

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Master of Arts

Troy L. Cox, Chair
Matthew P. Wilcox
Teresa R. Bell

Center for Language Studies
Brigham Young University

Copyright © 2022 Colin Lilya

All Rights Reserved

ABSTRACT

Examining the Willingness to Communicate (WTC) Scale with Advanced Foreign Language Learners

Colin Lilya

Center for Language Studies, BYU

Master of Arts

This study explored the effectiveness of a Willingness to Communicate (WTC) scale with advanced foreign language (L2) learners. The WTC instrument included sections on communicating with native speakers of the L2 and peer language learners in various settings (at home, abroad, in-class and online). As most prior research had focused on beginning language learners, for this study, we recruited participants who began to learn their respective languages in informal, long-term immersion settings. Participants took the WTC survey as part of a larger self-assessment instrument with a subset of 600 intermediate and advanced level Spanish ($n = 339$), Portuguese ($n = 155$), and French ($n = 106$) L2 students taking an Oral Proficiency Interview (computerized). The instrument was found to be reliable (Cronbach $\alpha = .88$), and there was a significant difference [$t(5) = 2.97, p = .031$] in WTC between sections for online and in-class settings. However, the WTC had no significant relationship (Pearson's $r^2 = .0005$) with OPIc score. Thus, while WTC might help beginning learners reach advanced level language, it might not discriminate among learners who are already advanced.

Keywords: Willingness to Communicate, oral proficiency, OPIc, advanced learners, immersion, self assessment

ACKNOWLEDGMENTS

Considering the shortened length of my time in the Second Language Teaching program here at Brigham Young University, I never would have finished all of my research and writing without the help and support of so many family members, friends, and faculty. I have been blessed to have so many family members and friends in the area to give me a refuge from studies, research, and writing when I needed it most. I am also grateful for my friends and peers, especially those who participated in the weekly writing and revision sessions where I learned an immeasurable amount about becoming a better writer and researcher. If not for that group, I doubt I ever would have created a final product of which I could be so proud, and I certainly would not have enjoyed the journey to this final destination as much as I have.

Most importantly, words cannot express enough thanks to my thesis chair, Dr. Troy Cox. Not only did he hold me accountable through the entire thesis process from start to finish, but he did so while teaching me invaluable lessons about patience, leadership, and scholarship along every step of the way. I could not have asked for a better mentor, and I will forever remember how much he truly cared about and served those around him. He is indeed a living embodiment of the prominent sign at the southwest entrance to our campus that reads, “Enter to Learn; Go Forth to Serve.”

Table of Contents

ABSTRACT	ii
ACKNOWLEDGMENTS.....	iii
Table of Contents.....	iv
List of Tables	vii
List of Figures.....	viii
Introduction	1
Literature Review.....	2
L1 WTC.....	2
WTC and SLA Theories.....	3
L2 WTC.....	4
L2 WTC Variables	4
Trait-based Variables.....	5
Situation-based Variables	6
Language Proficiency Measurements	8
WTC & Proficiency	9
Research Questions	9
Methodology.....	10
Participants	10
Instruments	11

WTC Instrument.....	11
ACTFL OPIc.....	13
Procedure	14
Survey and Test Administration.....	14
Data Analysis	14
Results	15
Rating Scale Calibration.....	15
Research Question 1: Instrument Reliability.....	16
Research Question 2: Differences in Difficulties Between WTC Instrument Subcategories ...	18
Research Question 3: Relationship between WTC and L2 Oral Proficiency.....	19
Relationship Without Extreme Persons.....	22
Discussion.....	23
Instrument Reliability.....	23
Observed vs. Measured WTC	25
Item Difficulty Between Sections	25
WTC and Proficiency.....	26
Proficiency Level	26
Limitations.....	27
Extreme Persons in the Data.....	27
Implications for Future Research.....	28

References	30
Appendix A: WTC Questionnaire	34
Willingness to Communicate in a Foreign-Language Scale (WTC-FLS)	34
Section 1: Speaking [Language] with Native [Language] -Speakers In My Home Country	34
Section 2: Communicating in [Language] with Native [Language]-Speakers Online	35
Section 3: Speaking [Language] with English-speakers in a School Setting In My Home Country	35
Section 4: Speaking [Language] with Native English-Speakers learning [Language] in the Mission Field	36

List of Tables

Table 1 <i>WTC Instrument Items</i>	12
Table 2 <i>WTC Instrument Sections Item Subtotals</i>	18
Table 3 <i>Welch 2-sided t-test between WTC Instrument Sections</i>	19
Table 4 <i>OPIc Ratings Person Subtotals</i>	20
Table 5 <i>Welch 2-sided t-test of WTC Logit Values Between OPIc Score Groups</i>	21
Table 6 <i>Welch 2-sided t-test of WTC Logit Values Between OPIc Score Groups (non-Extreme)</i>	23

List of Figures

Figure 1 <i>Stacked Column Histogram of Language Proficiency for Each Language</i>	11
Figure 2 <i>Andrich Thresholds for 4-Point Scale</i>	16
Figure 3 <i>Wright Map of Person Ability and Item Difficulty</i>	17
Figure 4 <i>Scatterplot of WTC Person Ability Measure and OPIc Score</i>	22

Introduction

Hardly a day goes by where we do not communicate with another person in one form or another, whether that be in person, over the phone, via direct messaging, or in writing, but the frequency and volume of communication may vary drastically by the circumstances or by the individual. One construct that helps to define and measure these differences is called Willingness to Communicate (WTC; McCroskey & Baer, 1985). McCroskey and Baer originally defined WTC in the field of first language (L1) acquisition as a personality trait that dictates the degree of predisposition to consistently communicate across a variety of different situations. Second language (L2) acquisition researchers posited that this trait could also be a contributing factor to success for L2 language learners (MacIntyre et al, 1998).

However, in contrast to the static model of WTC created by McCroskey and Baer (1985) for L1 contexts, MacIntyre et al. (1998) constructed a situational model of dynamic variables that may influence WTC over the course of even just a single conversation. This model was later affirmed by several other researchers (Kang, 2005; S & Hsieh, 2019; Lee & Lee, 2020; MacIntyre & Legatto, 2011), and is the basis for most of the WTC instruments in use today. While WTC is a complex construct that is influenced by many trait- and situation-based variables, two particularly influential components determine an individual's WTC: the nature of the interlocutor (Kang, 2005; MacIntyre, 1998; Sato, 2020) and the contextual setting under which the conversation takes place (Lee & Drajiati, 2020; Lee & Lee, 2020).

Though many studies have sought to identify the variables that affect WTC, fewer studies have made efforts to determine the extent to which a WTC measure can predict L2 oral language proficiency. This previous research was largely composed of learners of English as a second language in formal learning environments (Al-Murtadha, 2021; Darasawang & Reinders, 2021;

Liu & Jackson, 2008; Robson, 2015; Sato, 2020). Thus, research on the degree of correlation between the WTC and oral language proficiency of advanced learners of non-English second languages in long-term, informal learning environments is scant. The purpose of this thesis, then, is to attempt to address this gap in the research and further examine whether learners' WTC in different contexts of communication might better predict oral language proficiency.

Literature Review

L1 WTC

McCroskey and Baer (1985) argued that WTC is a key trait for developing meaningful interpersonal relationships because it contributes to participation in social learning. This idea is rooted in Vygotsky's (1978) concept of Sociocultural Theory, which claims that linguistic, cognitive, and social learning development occurs under collaborative, sociocultural contexts. In L1 research, WTC finds its origins in ideas such as unwillingness to communicate, verbal behavior tendencies, and shyness.

Though initial research on communication behaviors primarily focused on personality variables like introversion and self-esteem as indirect predictors of communication outcomes, Burgoon and Burgoon (1974) created a new scale to concentrate on actual, rather than indirect, predictors. Items on this instrument included levels of agreement with statements such as "I'm afraid to speak up in conversations" and "I find it easy to make conversation with strangers" (p. 64). Burgoon (1976) later refined this scale to clarify weak items and items that did not distinguish between the two underlying factors that had been identified, which were communication distrust and apprehension.

Later shifts in perspective in the field of communication turned attention from unwillingness to communicate toward its inverse, WTC. McCroskey and Baer (1985) defined

WTC in the L1 and were the first to develop a scale to measure WTC. McCroskey (1992) later confirmed the construct validity of WTC, indicating that it contains an appropriate level of conceptual distinctiveness from related constructs, such as extroversion. Following studies further validated the instrument for university students among groups of strangers, acquaintances, and friends (Asker, 1998; McCroskey & Baer, 1985).

WTC and SLA Theories

Research in L1 WTC in creating interpersonal relationships led to applications in L2 environments, where both the purpose and process of learning often include communication with others. These applications are concretely based in language acquisition theories. Some of the most applicable theories include Swain's Output Hypothesis (Swain, 1993), which posits that along with comprehensible input, opportunities for output by learners in the L2 is necessary for language development, and Long's Interaction Hypothesis (Long, 1981), which adds extralinguistic features and negotiation of meaning to the list of necessary experiences. But perhaps the most relevant theory is sociocultural theory, which asserts that linguistic competency is constructed through social and cultural interactions between experts and novices in the learner's Zone of Proximal Development (Vygotsky, 1978).

The Zone of Proximal Development is the theoretical space above a learner's individual capabilities where they may perform with the additional help of an expert. Interestingly, the expert need not be a native speaker, but instead peers may serve just as effectively as relative experts in some areas where they may help each other co-construct meaning to perform linguistic tasks (Donato, 1994). Assuming this is true, it would therefore make sense that higher L2 WTC – and thus more meaningful interactions – would lead to greater development in language skills.

L2 WTC

Thus, although WTC was originally researched in L1 communication communities, it has since expanded to include how willing learners are to speak in L2 contexts. In fact, Al-Murtadha (2021) found moderate, positive relationships between L1 and L2 WTC in a semester-long study with English as a foreign language (EFL) secondary school students in Yemen. Several researchers also explored the situational nature of L2 WTC and conducted studies that uncovered some of the underlying variables, which will be further discussed below. These studies established that WTC changes situationally throughout time and even over the course of a single conversation.

L2 WTC Variables

When WTC research first started in L1 contexts, McCroskey and Baer (1985) suggested that WTC is consistent across all contexts for an individual in comparison to others, but acknowledge that, based on their findings, “the larger the number of receivers and the more distant the relationship of the individual with the receiver(s), the less willing the individual is to communicate” (p. 8). However, they did not fully address the magnitude of impact of these situational variables. In later studies, MacIntyre et al.’s (1998) model and adaption of WTC to L2 contexts resulted in the discovery of over 30 variables that influence WTC. These variables indicated a need to change some elements of continuity in the definition of WTC, including elements of transience and situation, for which were otherwise unaccounted. Each of these variables existed on a continuum from the permanent, trait-based variables that are more derivative of the individual’s personality to more temporary, situational variables that may affect a speaker’s WTC.

Trait-based Variables

A combination of early research in WTC and more recent additions to the field has yielded several different possible underlying traits that help predict WTC. In McCroskey and Baer's (1985) study, they found that communication apprehension may be the strongest predictor of WTC, but that other factors are likely to also have a significant impact. They noted that situational variables are also likely to have an impact, but that individual trends would remain consistent. For example, if one speaker has a higher WTC than another speaker in one context, then McCroskey and Baer asserted that they should also have a higher WTC in other contexts as well. MacIntyre et al. (1998) similarly attributed the more permanent differences between individuals' WTC to such factors as communication behavior and motivational inclinations.

A study by Lee and Hsieh (2019) showed across three separate contexts that L2-specific self-confidence was a significant predictor of L2 WTC, which further validates the previous studies, simply in the inverse. The relationship between L2 self-confidence and WTC was also validated by Sato (2020). Lee and Hsieh (2019) also found that grit, which is defined as long term persistence and enthusiasm towards goals, has a positive correlation with L2 WTC. Another particularly intriguing finding from Lee and Hsieh's study is that L2-specific anxiety did not negatively correlate with WTC in digital settings, which matches the findings of other research that L2 anxiety tends to be lower in digital settings.

In their later article, Lee and Lee (2020) also found that grit ($r = .32$) and motivation ($r = .56$) both revealed strong correlations with WTC in the classroom, as had been shown in beforementioned studies. However, grit did not appear to be a significant ($p < .05$) predictor of WTC in digital settings, where self-confidence ($r = .49$), risk-taking ($r = .22$), and virtual intercultural experiences ($r = .51$) manifested more strongly.

These findings show that it is likely that the higher L2 self-confidence (and inversely, lower communication apprehension) manifests in increased attempts to interact with other speakers of the target language (TL), and that higher grit manifests itself in greater desire to achieve the long-term goal of language proficiency and thus a desire to speak with other TL speakers to practice and work toward their goal. However, as Lee and Lee (2020) uncovered in their study, the extent to which some of these traits influence a speaker's WTC can be highly dependent on the context.

Situation-based Variables

In several studies of trait-based variables influencing WTC, researchers have revealed a few situationally transient factors. Some research even suggests that WTC may fluctuate over the course of a single interaction. For example, MacIntyre and Legatto (2011) found that WTC is a dynamic system where each state was influenced by the preceding state and influenced the following in turn. In other words, the same person in the same situation may have different levels of WTC based on the previous interactions that they experience. Kang (2005) further emphasized the situational nature of L2 WTC and found that WTC changed situationally throughout time and confirmed its fluctuation over the course of one conversation. In fact, in a study by Sato (2020), it was found that all speakers' WTC fluctuated throughout the session, regardless of trait-level WTC found before the study.

Interlocutors. As McCroskey and Baer found in their study, the number and nature of the interlocutors may have a strong influence on L2 WTC. Specifically, they found that the larger the number of interlocutors and the more distant of a relationship they have with the speaker, the less likely the individual is to engage in communication. This is confirmed by MacIntyre et al. (1998), who identified the social and individual context as correlated variables.

Specifically, they mentioned that physical distance, physical attraction, frequency of contact, and similarity between speakers as the primary elements of desire to communicate with a specific person.

Kang (2005) also conducted a qualitative study to identify key factors that influenced WTC moment to moment within a single conversation and found feelings of security, excitement, and responsibility to be the most impactful, each of which was impacted by their conversation partners. For example, partner familiarity and number of interlocutors contributed to feelings of security, which were requisite for lower communication apprehension. Sato (2020) also identified a significant influence of interlocutors on sense of security. Furthermore, the nature of the interlocutors (native/non-native speakers), interlocutors' appearance (one participant felt uncomfortable about a tongue piercing), and perceived interest of the interlocutor on the topic being discussed all affected feelings of excitement in Kang's (2005) study. Lastly, the greater the number of people in the conversation, the less responsible everyone felt for contributing to the discussion.

Context. After the role of interlocutors, the context of the speaking opportunity had the next most common influence on L2 WTC. Though "context" can be interpreted broadly, general themes such as affective-cognitive context, topic, and physical and temporal location will suffice for the purposes of this paper.

Some of the variables identified in the research conducted by MacIntyre et al. (1998) included situated antecedents and affective-cognitive context. One layer of the situated antecedents includes state communicative self-confidence, which refers to a speaker's perceived confidence in the topic of discussion and lack of anxiety. Furthermore, the affective-cognitive context is influenced by the speaker's motivation to learn the language, their integration into the

cultural group, and the specific social situation. Many of these ideas were later confirmed by other researchers.

For example, Kang (2005) identified topic familiarity and conversational context (how far along the interlocutors are in the conversation) as two factors influencing a sense of security among Korean EFL students. The amount of interest that an individual has in a topic also contributed to feelings of excitement, promoting speaker L2 WTC, which was also found in Sato's (2020) qualitative study of four [Language] EFL students. Specifically, Sato found that opportunities to talk about oneself and one's opinions had a more pronounced influence on the advanced speakers' WTC. Moreover, the primary factors identified that influenced feelings of responsibility were perceived topic importance, greater understanding of a topic than their interlocutors, and defense of their culture regarding sensitive issues (Kang, 2005).

In Lee and Lee's (2020) study of Korean undergraduate and graduate EFL students WTC was found to be highest outside the classroom, then in digital settings, and finally inside the classroom. Specifically, the younger students who had more exposure to cultural experiences through technology had the highest L2 WTC in digital settings. Lee and Drajati (2020) also researched digital WTC and created a revised WTC instrument to measure differences between WTC inside of classrooms, outside of classrooms, and in digital contexts for Indonesian EFL students. Their exploratory factor analysis indicated that these three factors are indeed distinct and showed that their scale had potential for outside application to other populations for further review.

Language Proficiency Measurements

Language proficiency can be measured in many different ways, but this research focuses on oral mode as measured by ACTFL's Oral Proficiency Interview – Computer (OPIc; ACTFL,

2018). Prior to the creation of the OPIc, ACTFL's Oral Proficiency Interview (OPI) was one of the primary assessments of oral proficiency. Despite early problems with reliability among raters of this evaluation, improved training by ACTFL yielded much higher consistency and reliability between raters (Surface & Dierdorff, 2003; Surface & Dierdorff, 2008). Thompson et al. (2016) later found the OPIc to be a comparable method of measuring learners' proficiency to the OPI, and though the scores were slightly higher than the OPI, the effect size was small enough that the difference was minimal.

WTC & Proficiency

Previous research on the effects of WTC on oral and written language proficiency is mixed. Some of this research is inconclusive about the effects of WTC on oral proficiency, if any at all (Robson, 2015), though it does mention the influence of individuals' personal culture on their WTC. In contrast, some research successfully identified a positive correlation between observed L2 WTC and proficiency (Al-Murtadha, 2021; Darasawang & Reinders, 2021; Sato, 2020). Each of these studies examined EFL speakers over a spectrum of proficiencies who experienced most of their language learning in formal learning environments. There is also a large amount of research that outlines the definitions of WTC and the variables that influence the long-term, trait-like characteristics and short-term situational factors that influence it.

Research Questions

Though Al-Murtadha (2021) and Liu and Jackson (2008) both found the WTC survey to be reliable (Cronbach $\alpha > .8$) in studies correlating WTC and L2 oral proficiency, reliability for the population of the present study or for WTC instruments that include online WTC were unknown. Because the previous research does not look at long-term studies of advanced speakers of non-English second languages across different language domains, there is a need to research

this population with an L2-specific WTC instrument. Thus, the following research questions will be addressed in the present study:

1. To what extent does the WTC instrument function reliably for this population?
2. How do the four subcategories of the WTC instrument vary?
3. What is the relationship between WTC and L2 oral proficiency?

Methodology

The population of interest will be discussed in the Participants section, and the specific WTC instrument and its subcategories will be discussed further in the Instruments section, which will also outline the methods of L2 proficiency measurement.

Participants

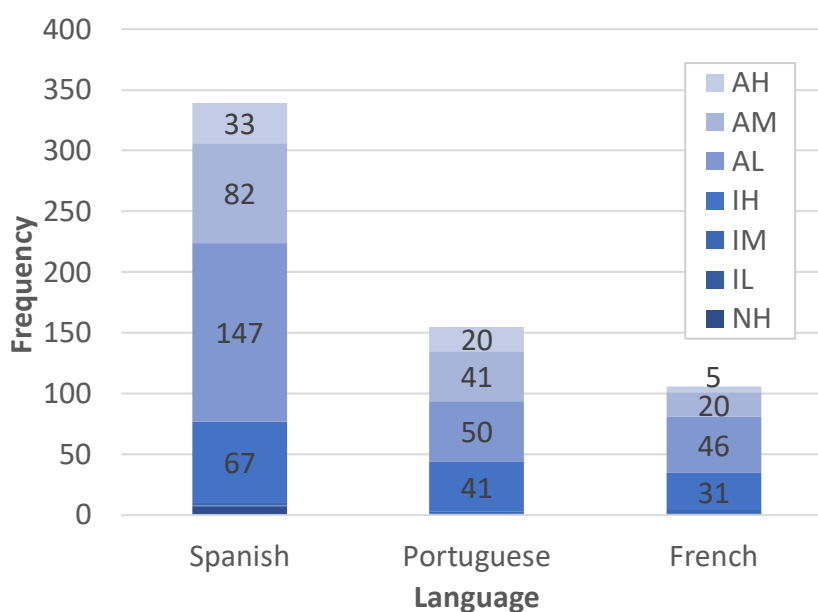
At this university, most of the language learners in intermediate and advanced language courses learned their respective L2's during missionary service ($n = 539$) rather than in more traditional learning environments. While these immersion experiences typically lasted from 18 to 24 months, many of the participants in this study spent less than this due to COVID-19 restrictions. One key aspect of the participants' immersion environments was a religious imperative to "open your mouth." This refers to encouragement from during their initial missionary training to always seek out opportunities to talk with people in the areas where they lived.

The participants were studying Spanish ($n = 339$), Portuguese ($n = 155$), and French ($n = 106$), and were enrolled in 300-level university language courses. Over 97% of the participants had OPIc ratings of Intermediate High or higher, but there were 17 whose scores ranged from Novice High to Intermediate Mid. Figure 1 shows the specific distribution of languages and

proficiency levels. In instances when students took the instrument multiple times, only the data from their first completed survey was used.

Figure 1

Stacked Column Histogram of Language Proficiency for Each Language



Note. The abbreviations in the legend refer to ACTFL's proficiency levels in descending order: Advanced High, Advanced Mid, Advanced Low, Intermediate High, Intermediate Mid, Intermediate Low, and Novice High.

Instruments

Two primary instruments were used for this study: the WTC section of the Language Ability Self Evaluation Resource (LASER; Cox, 2022) and ACTFL's OPIc (ACTFL, 2018).

WTC Instrument

The WTC is a component of the LASER, a 30-minute self-assessment questionnaire which covers a variety of different topics, including basic learner information, language learning background, WTC, and Grit. The questions in the LASER were mostly adapted from a previous

research instrument (Baghaei, 2013), which had 3 subcategories. As online communication had emerged as a common setting in which language speakers engage, we decided to add an additional subcategory that addressed online communication with native speakers. Thus, the final instrument had 26 questions divided into 4 subcategories, of which participants responded to a subset of 17 questions:

- Section 1—Speaking with native speakers in the participant’s home country (k = 6 of 6)
- Section 2 (new)—Communicating with native speakers online (k = 4 of 7)
- Section 3—Speaking with L2 TL speakers in a school setting (k = 3 of 5)
- Section 4—Speaking with L2 TL speakers abroad (k = 4 of 8)

A list of all items included in the instrument can be found in Table 1, and a more in-depth description can be found in Appendix A. Participants responded to each item using a sliding scale with values from 1 to 100 to indicate the percentage of time that the statements describe them.

Table 1

WTC Instrument Items

Item	Statement
1.1	If I encountered some native speakers of [Language] in the street, restaurant, hotel, etc., I hope an opportunity would arise and they would talk to me.
1.2	If I encountered some native speakers of [Language] in the street, restaurant, hotel, etc., I would find an excuse and would talk to them.
1.3	If I encountered some native speakers of [Language] who are facing problems in my country because of not knowing our language, I would take advantage of this opportunity and would talk to them.
1.4	I am willing to accompany some native speakers of [Language] and be their tour guide for a day free of charge.
1.5	I am willing to talk with native speakers of [Language].
1.6	If someone introduced me to a native-speaker of [Language] I would like to try my abilities in communicating with him/her in [Language].
2.1	If I encountered some native speakers of [Language] in an online forum, I hope an opportunity would arise and they would message me.
2.2	If I encountered some native speakers of [Language] in an online forum, I would find an excuse and would message them.

Item	Statement
2.3	If I encountered some native speakers of [Language] who are facing problems online because of not knowing our language, I would take advantage of this opportunity and would message them.
2.4	I am willing to communicate with native speakers of [Language] online.
3.1	In order to practice my [Language], I am willing to talk in [Language] with my classmates outside the class.
3.2	I am willing to ask questions in [Language] in the classes at the university.
3.3	I am willing to talk and express my opinions in [Language] in the class when all my classmates are listening to me.
3.4	I am willing to have pair and group activities in the class so that I can talk in [Language] with my classmates.
3.5	In order to practice my [Language] I am willing to talk in [Language] with my professors outside the class.
3.6	I am willing to give a presentation in [Language] in front of my classmates.
3.7	In group work activities in the class when the group is composed of my friends, I am willing to speak in [Language].
3.8	In group work activities in the class when the group is NOT composed of my friends, I am willing to speak in [Language].
4.1	In order to practice my [Language], I am willing to speak in [Language] with native English-speaking missionaries learning [Language].
4.2	I am willing to ask questions in [Language] in large group settings of other native English-speaking missionaries learning [Language].
4.3	I am willing to talk and express my opinions in [Language] in large group settings of other native English-speaking missionaries learning [Language] that are listening to me.
4.4	I am willing to speak in [Language] in pair and small group settings with other native English-speaking missionaries learning [Language].
4.5	I am willing to teach a lesson in [Language] in front of a large group of other native English-speaking missionaries learning [Language].
4.6	I am willing to speak in [Language] with a group composed of friends that are also [Language] learners.
4.7	I am willing to speak in [Language] with a group NOT composed of friends that are also [Language] learners.

ACTFL OPIc

The ACTFL OPIc, much like the OPI, is a test which provides oral proficiency evaluations (ACTFL, 2018). It is delivered through the internet, and prompts are given through a computer program so that the test can be taken at the candidate's convenience. The format consists of an introduction, background survey, self-assessment, and forms section. The answers from the background survey are used by the computer algorithm to randomly select topics for later tasks. Then, the self-assessment results determine which of the five test forms the computer

will use, which identifies the proficiency range in which the candidate will be rated. Certified raters later review the interview for function, accuracy, content/context, and text type in order to assign a proficiency score according to the ACTFL proficiency guidelines (ACTFL, 2012).

Procedure

Survey and Test Administration

The self-report module of the questionnaire was administered online for participants through a Qualtrics survey, so they had the freedom to conduct the survey in their own time and setting. Instructors sent the participants a link to the LASER at the beginning of the semester and asked students to complete it within a few weeks. WTC was only one section of this self-report module and was predicted to take about eight minutes to complete. Following completion of the self-report module, participants were emailed links to register for and take the ACTFL OPIc within one month. Procedures for administration of the OPIc can be found on ACTFL's official website (ACTFL, 2018).

Data Analysis

Quantitative measurement and analysis of WTC data gathered from the LASER and language proficiency data from the OPIc tests provided insight into the degree to which they are correlated. Large numbers of participants in the survey study helped provide statistical power to identify the effect that WTC had as one of many variables that influences proficiency.

The data analysis was performed using Rasch analysis in Winsteps. Because the survey uses sliding scales to gather participants' responses, the data is considered polytomous and was analyzed using the rating scales model. Participants were labelled with their OPIc rating. To answer the first research question, the rating scale was analyzed and calibrated to find the best number of rating categories, and then the internal reliability was calculated using Cronbach

alpha. To answer the second research question, the item difficulty parameters were evaluated based on the 4 sections they belonged to. T-tests (Welch's) were then used to compare if there were systematic differences in the likelihood of a participant to endorse the different sections. To answer the third question, we examined the relationship between the Rasch person ability and the OPIc rating. We further used t-tests (Welch's) to examine any systematic differences among the ratings.

Results

After some rating scale calibration, the WTC data from the LASER provided information about the reliability of the instrument, the extent to which the four subcategories of the WTC instrument varied in difficulty or likelihood to endorse, and the relationship between WTC and L2 oral proficiency.

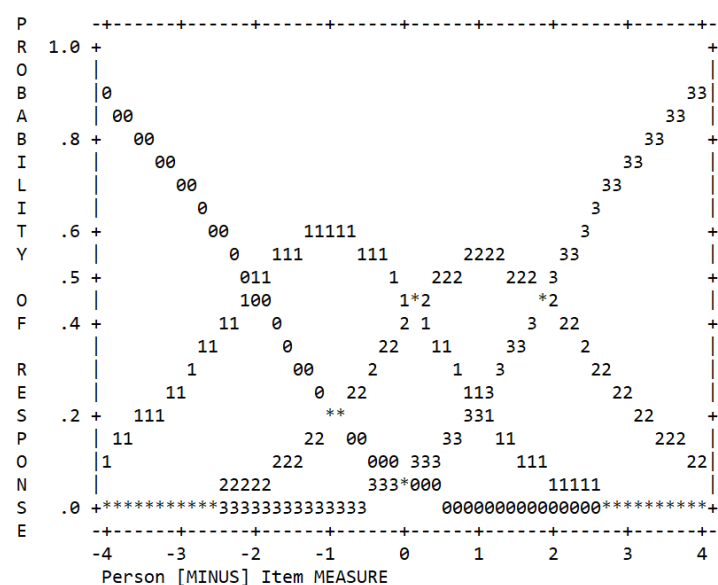
Rating Scale Calibration

Initial data analysis with the full 100-point scale for each WTC item resulted in disordered Andrich thresholds despite good reliability ($\alpha = .85$). Disordered Andrich thresholds are indicative of narrow categories that measure the latent variable – which in this case is WTC – and tend to have a poor fit between the model and the data (Linacre, 2018). Simply put, poorly ordered or narrow categories do not evenly categorize responses that fall between two categories. For example, if 30 responses fell between a 2 and a 3, properly ordered categories would result in observations being split evenly between a 2 and a 3, whereas disordered thresholds would result in observations favoring one or the other. Linacre (2018) claimed that disordered categories such as these can occur when raters cannot discriminate between the categories from which they are choosing on an item.

Thus, the rating scale was adjusted to a 10-point scale, which improved the category probabilities and reliability ($\alpha = .91$), but a final adjustment to a 4-point scale resulted in the most acceptable thresholds (see Figure 2) despite a slightly lower reliability ($\alpha = .88$). This was accomplished by rounding response values down to the nearest 10 and dividing by 10 (i.e., 9 to 0 and 34 to 3). Responses of 100 were rescaled to 9's. Then, the scores were adjusted from 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 to 0, 0, 1, 1, 1, 1, 2, 2, 2, and 3, respectively.

Figure 2

Andrich Thresholds for 4-Point Scale

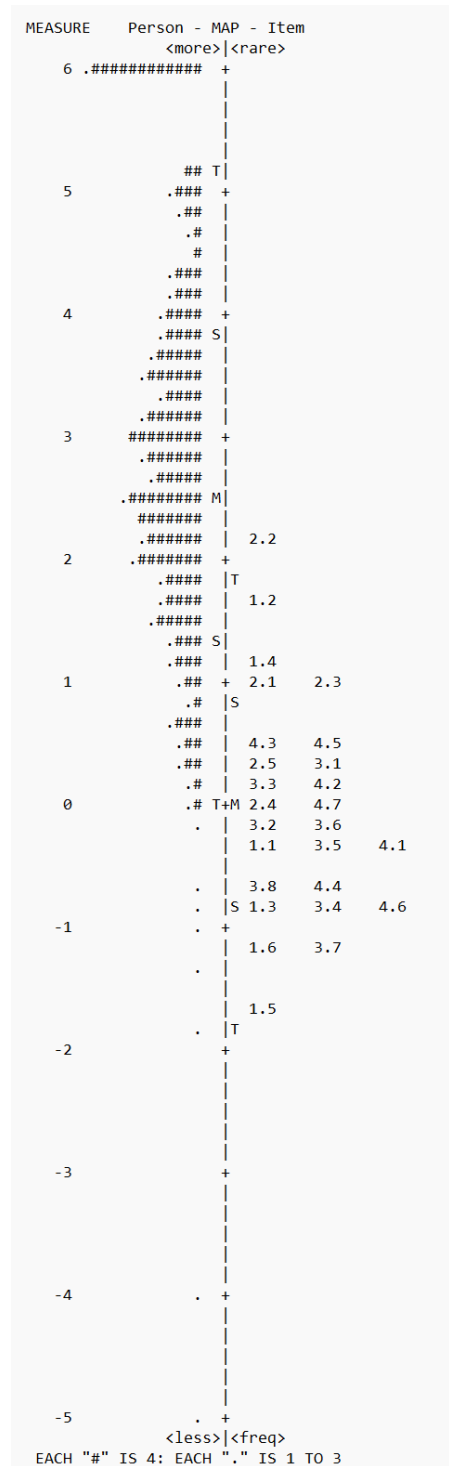


Research Question 1: Instrument Reliability

The WTC section of the LASER was found to be internally consistent, with a Cronbach alpha of 0.88. However, as can be seen in Figure 3, the distribution of the measured person ability and item difficulty measures were not congruent. In this context, the item difficulty refers to participants' likelihood to endorse the statement in that specific item.

Figure 3

Wright Map of Person Ability and Item Difficulty



In Figure 3, the left column shows the distribution of persons, and the right column shows the distribution of items. Notice that the mean of the persons, signified by “M,” is around 2.5 whereas the mean of the items, also signified by “M,” is centered at 0 as designated by the parameters in the Rasch model used. This shows that participants were much more likely to positively endorse WTC statements than to not endorse them. Also notice that the mode ($n = 51$) is at the maximum value, indicating that there are many participants who indicated that they were willing and able to “open your mouths” at all times.

Research Question 2: Differences in Difficulties Between WTC Instrument Subcategories

In response to the second research question, the different sections within the WTC instrument were evaluated to determine the differences in item difficulty. The most difficult section was speaking with native speakers online (Section 2, see Table 2), which had an average of .9, meaning that it was less common for participants to engage in communication online. The easiest was speaking with L2 TL speakers in a school setting (Section 3).

Table 2

WTC Instrument Sections Item Subtotals

Section	Count	Score	Measure	Mean	SD	Outfit
*	26	854	0	0.18	0.87	0.93
1–Native	6	1511	-0.17	0.54	1.2	0.99
2–Online	5	806.2	0.90	0.38	0.75	0.99
3–School	8	445.4	-0.34	0.18	0.44	0.84
4–L2 Abroad	7	791.9	-0.11	0.20	0.48	0.92

An ANOVA [$F(3,25) = 2.54, p = .08$] did not find a significant difference overall, however, we did find a significant difference between communicating with native speakers online and speaking with L2 TL speakers in a school setting [$t(5) = 2.97, p = .031$]. A summary of these statistics is found in Table 3.

Table 3

Welch 2-sided t-test between WTC Instrument Sections

Section	Section	Mean Difference Measure of WTC	SE	t	df	Welch 2-sided p
1—Native	2—Online	-1.07	0.66	-1.63	8	0.143
1—Native	3—School	0.17	0.57	0.30	6	0.775
1—Native	4—L2 Abroad	-0.07	0.57	-0.11	6	0.913
2—Online	3—School	1.23	0.42	2.97	5	0.031
2—Online	4—L2 Abroad	1.00	0.43	2.35	6	0.057
3—School	4—L2 Abroad	-0.23	0.27	-0.87	12	0.400

Research Question 3: Relationship between WTC and L2 Oral Proficiency

In response to the third research question, the different oral proficiency ratings were evaluated to determine the differences in WTC. Because too few participants received OPIc ratings of Novice High (n = 8), Intermediate Low (n = 2), and Intermediate Mid (n = 7), their results will be excluded from further discussion. The OPIc rating with the highest WTC was Intermediate High (see Table 4), which had an average of 3.05, and the rating with the lowest WTC was Advanced Low, which had an average of 2.7.

Table 4*OPIc Ratings Person Subtotals*

Section	With Extreme Persons				Without Extreme Persons			
	Count	Score	Measure	<i>SD</i>	Count	Score	Measure	<i>SD</i>
*	600	37	2.86	1.52	547	36.5	2.58	1.19
NH	8	40.4	2.65	0.62	8	40.4	2.65	0.62
IL	2	32.5	2.46	0	2	32.5	2.46	0
IM	7	39.4	3.99	0.72	6	39.5	3.6	0.31
IH	139	37.9	3.05	1.33	128	37.3	2.77	1.12
AL	243	36.4	2.70	1.72	221	36	2.46	1.31
AM	143	37.1	2.94	1.37	128	36.4	2.55	1.05
AH	58	36.5	2.84	1.42	54	36.1	2.59	1.25

Note. The abbreviations in the legend refer to ACTFL's proficiency levels in ascending order: Novice High, Intermediate Low, Intermediate High, Advanced Low, Advanced Mid, and Advanced High.

An ANOVA [$F(6, 599) = 1.22, p = .29$] did not find a significant difference overall, and we found no significant differences between any of the OPIc ratings Person Ability Measures. The Person Ability Measures were calculated estimations of each participant's WTC based on their responses on all the items. Welch's 2-sided probability t-tests between proficiency scores did not show any significant ($p < .05$) differences between the WTC of the participants; thus, the data fails to reject the null hypothesis (H_0 : there is no significant difference between the WTC of participants at different OPIc levels). A summary of these statistics is found in Table 5.

Table 5*Welch 2-sided t-test of WTC Logit Values Between OPIc Score Groups*

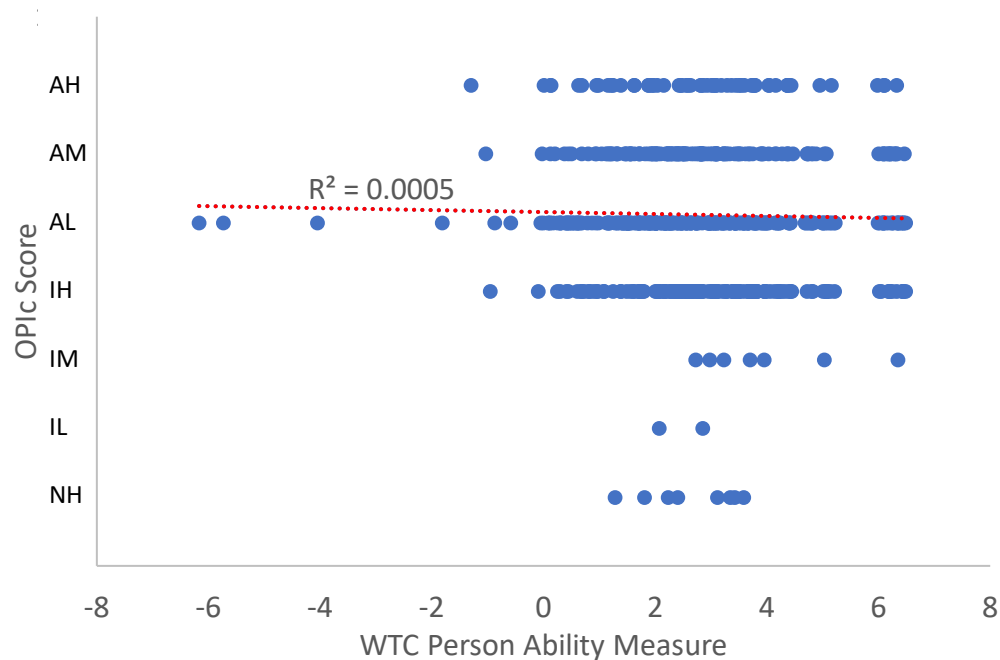
OPIc Rating	OPIc Rating	Mean Difference Measure of WTC	SE	t	df	Welch 2-sided p
IH	AL	0.34	0.18	1.91	335	.057
IH	AM	0.11	0.19	0.59	279	.558
IH	AH	0.21	0.25	0.84	102	.403
AL	AM	-0.23	0.18	-1.29	338	.199
AL	AH	-0.13	0.24	-0.54	98	.590
AM	AH	0.10	0.25	0.40	104	.691

Note. The abbreviations in the legend refer to ACTFL's proficiency levels in ascending order: Intermediate High, Advanced Low, Advanced Mid, and Advanced High.

A Pearson product-moment correlation [$r(599) = -.022, p < .05$] of WTC and OPIc scores also showed no significant correlation between the variables, which confirmed the Welch's t-tests. A scatterplot of Person Ability and OPIc score is shown in Figure 4, which shows each participant's Person Ability Measure versus their OPIc score. OPIc scores are represented with abbreviations for each sublevel, matching those of Figure 1.

Figure 4

Scatterplot of WTC Person Ability Measure and OPIc Score



Relationship Without Extreme Persons

When participants are extreme, for instance they always select the maximum value, it could be an instance that the participants did not read the statements and merely selected the same response category for all items. Thus, we did the analysis a second time excluding the extreme persons. We found a slightly stronger relationship between WTC and oral language proficiency. Of the 53 extreme persons, 51 had maximum self-reported values for WTC and the remaining 2 had very low self-reported WTC. Excluding these extremes, a Welch's 2-sided probability t-tests between proficiency scores showed a significant difference in WTC between participants with Intermediate High and Advanced Low scores [$t(291) = 2.07, p = .039$]. That is, the IH had a significantly higher WTC score than the AL with a mean difference of .31 logits. A summary of these statistics is found in Table 6.

Table 6*Welch 2-sided t-test of WTC Logit Values Between OPIc Score Groups (non-Extreme)*

OPIc Rating	OPIc Rating	Mean Difference Measure of WTC	SE	t	df	Welch 2-sided p
IH	AL	0.31	0.15	2.07	291	0.039
IH	AM	0.22	0.15	1.4	252	0.163
IH	AH	0.18	0.22	0.8	92	0.425
AL	AM	-0.09	0.14	-0.64	305	0.524
AL	AH	-0.13	0.21	-0.62	82	0.539
AM	AH	-0.04	0.22	-0.18	87	0.854

Discussion

The purpose of this study was to explore the relationship between WTC and oral language proficiency. In order to accomplish this, the first set of analyses confirmed the reliability of the LASER's WTC section as an instrument, the second explored the relationship between WTC and OPIc scores, and the third examined how the different sections of the WTC instrument interacted to understand what aspects may have weaker or stronger influences on proficiency. Although the LASER was found to be reliable, no significant relationship was found between WTC measures and language proficiency. However, out of all the sections of the WTC instrument, the section about speaking with native speakers of the TL showed significantly different results from the other three.

Instrument Reliability

Our version of the WTC instrument, which was adapted from Baghaei (2013), was reliable, but the need to reduce the number of response categories provided interesting feedback that can be applied to future studies. It is difficult to distinguish between the difference between being willing to communicate 76% of the time versus 77% of the time, so a 100-point scale may

be less helpful than a smaller, 4-point scale. However, because the WTC instrument was largely adapted from the work of Baghei (2013), it was expected that it would still be reliable for this population, and the present study confirms this hypothesis. Furthermore, the domain-specific WTC sections were shown to be reliable, so the new digital section can be used for replication and future studies, which grows ever relevant given the increased use of technology both in daily life and in education.

Responses to the WTC instrument demonstrated sufficient reliability and distinction given the Cronbach's alpha value ($\alpha = .88$) but overall demonstrated very high measures of WTC. However, no previous research indicated any mismatch between person ability and likelihood to endorse with respect to participants responses to the WTC instrument, as was seen in this study. Linacre (2002) stated that outfit statistics are more sensitive to responses in which the difficulty of the item and person ability have large separation, which helps draw a relationship between the large number of misfit participants and mismatch between item difficulty and person ability. Among the possible explanations for this incongruence, the cumbersome 100-point scale (which has already been discussed) and the specific context under which the participants studied their respective languages provide the most clarity. Several previous studies (Kang, 2005; Lee & Drajiati, 2020; Lee & Lee, 2020; MacIntyre et al., 1998; Sato, 2020) have shown that conversational context has a strong influence on participants' WTC, and the religious mission provides one such context which may encourage unusually high levels of WTC.

The findings of the present study specifically endorse that of the study by Kang (2005), which discovered that two primary factors that influenced a speaker's WTC were 1) perceived importance of conversation topic and 2) greater understanding of the conversation topic than their interlocutors. Certainly, young men and women who volunteered between a year and a half

to two years towards devoting themselves wholly to their religion perceived that their reasons to communicate were important and also developed strong understandings of the religious content of their conversations.

Observed vs. Measured WTC

Self-reported statistics are not always accurate indications of participants' true ability, and WTC is no exception. In fact, Al-Murtadha (2021) found that neither self-reported levels of L1 nor L2 WTC accurately reflected observed WTC in a classroom environment. While that study found that observed WTC was related to language proficiency in a semester-long study, there was no predictive value of self-reported WTC for language proficiency, which the present study further validates.

Item Difficulty Between Sections

The finding that the only significant ($\alpha < 0.05$) difference in likelihood to endorse WTC items between sections was between communicating with native speakers online (Section 2) and speaking with L2 TL speakers in a school setting (Section 3) contrasts the findings of several other studies which show higher digital WTC compared to other modes. For example, Lee and Lee's (2020) study found that WTC was consistently higher in digital settings compared to inside the classroom and that students with more cultural exposure had the highest L2 WTC in digital settings. Thus, it would have been logical that the participants of the present study, who spent long durations immersed in the target culture, to have followed the same pattern. However, it is unclear why our participants consistently demonstrated lower likelihood to endorse this digital WTC. This distinction between the two contexts of WTC reinforces theories of context-dependent and transient WTC compared to purely trait-based models.

WTC and Proficiency

Contrary to expectation based on sociocultural theory and anecdotal evidence, this study found no significant ($p < .05$) correlation between WTC and language proficiency. This lack of correlation stands in contrast to that of Robson (2015) but confirms the findings of several researchers (Darasawang & Reinders, 2021; Liu & Jackson, 2008; Al-Murtadha, 2021; Sato, 2020). However, the key similarities and differences between the present study and previous studies highlight two key relationships regarding WTC and proficiency, namely, participant proficiency level and method of WTC measurement.

Proficiency Level

Whereas the studies by Darasawang and Reinders (2021), Liu and Jackson (2008), and Sato (2020) found a significant, positive relationship between WTC and proficiency level compared to the lack of correlation in the present study, the overall higher proficiency level of the participants may explain the dissimilar findings. In particular, the Darasawang and Reinders (2021) and Liu and Jackson (2008) studies investigated the case of lower proficiency language users and found significant relationships. Thus, it may be the case that WTC is only a significant predictor of proficiency on the lower end of the proficiency spectrum and the participants in the present study were slightly beyond the useful range of application.

Furthermore, the findings from Sato (2020) showed that, in a qualitatively measured research design, the WTC of two advanced participants was higher than that of two intermediate participants. This does not necessarily conflict with the findings of the present study, which primarily observed advanced speakers. Because of low participation in the study of individuals with proficiency levels below Intermediate High, it cannot be claimed that there is no

relationship across larger spans of proficiency. Instead, WTC may have a floor effect in that it may not discriminate among once they have reached upper levels of proficiency.

Limitations

A few of the significant limitations of this study revolved around the unique population that the present study focuses on and the nature of self-reported data. From this study, 90% of the participants (n = 539) identified religious missions as the primary environment for their language learning. As has been mentioned, the former missionaries who participated in this study differ from traditional populations of students in that they spent most of their language learning in informal learning and immersion settings and that they have a unique directive to spend much of their time talking with native speakers of their TL as possible. Furthermore, self-reported measures of WTC may not be as accurate for this population because a large part of their view of personal value and success with the language may revolve around their perception of their own WTC in the TL, which they may not be able to easily separate from their missionary purpose. Lastly, low participation from individuals with novice or intermediate proficiencies limit the generalizability of these findings to mostly advanced speakers.

Extreme Persons in the Data

One significant limitation in this study is presence of the 53 extreme persons in the data. Most of these (n = 51) self-reported maximum WTC throughout the instrument, and there is no way of knowing whether this represents the participants' true opinions of their WTC or if their high scores are simply a result of not reading through and answering each item honestly. Disregarding their responses and assuming that they simply rushed through and maxed out every item, the data then shows that there is a significant difference between the WTC of Intermediate High and Advanced Low participants. Interestingly, the probabilities grew less and less

significant with each increase in proficiency level, as seen in Table 6. This may indicate that the influence of WTC diminishes as speakers increase their proficiency, but further research that looks at a range of proficiencies below advanced levels is necessary to provide evidence supporting this hypothesis.

Implications for Future Research

Additional research related to the topics of the present study could specifically improve in several ways: examining non-religious missionary populations, including larger numbers of participants across a larger range of proficiency levels, adjusting how WTC is measured, and focusing on WTC in more specific language domains.

Replications of this study with populations of advanced speakers who also spent time in extended immersion contexts would help to confirm the results of this study outside of the context of missionary volunteers who have such strong religious motivations. Where previous research has shown repeatedly that the context of both learning and interlocution have strong effects on WTC, it would be worthwhile to explore other areas.

Furthermore, both previous studies and the present study research specifically low or high proficiency level participants, with little that investigates wider ranges. Thus, it would be enlightening to pursue such research in order to examine more closely the relationship between WTC and language proficiency across longer periods of time and proficiencies to identify trends on a larger scale. Doing so may help distinguish at what point WTC no longer acts as a significant predictor of proficiency.

Additionally, research that focuses more on the manner of WTC measurement may help to clarify participants' self-assessment abilities and how they might be calibrated to reflect true values. Perhaps WTC may be more predictive of proficiency even at higher levels, as was seen in

the study by Al-Murtadha (2021), if participants could more accurately assess their own WTC. This kind of research is certainly resource intensive but may add significant value. One simple way to improve the present study would be to simply reduce the number of responses from which participants can choose. The data from this study required a similar process to reduce the range of responses in order to create more ordered Andrich Thresholds, but instead of giving participants the option to pick from a small number of choices that are easily distinguished may improve their ability to assess their individual WTC.

Lastly, because the digital language domain-specific WTC showed significantly lower likelihood to endorse, it may be worthwhile to further explore WTC in the different language domains and why they differ. There is other research on the different ways that WTC differs across domains (Lee & Drajati, 2020; Lee & Lee, 2020), but little to explain the possible causes of this phenomenon. Digital WTC itself is particularly relevant given the recent increases in online and remote learning, both as a function of recent global events and advances in communication and educational technology.

References

- ACTFL. (2012). *ACTFL proficiency guidelines 2012*. <https://www.actfl.org/resources/actfl-proficiency-guidelines-2012>
- ACTFL. (2018). *Oral proficiency interview–computer (OPIC)*. <https://www.actfl.org/center-assessment-research-and-development/actfl-assessments/actfl-postsecondary-assessments/oral-proficiency-interview-computer-opic>
- Al-Murtadha, M. A. (2021). The relationships among self-reported and observed first language and second language willingness to communicate and academic achievement. *Language, Culture and Curriculum*, 34(1), 80-94. <https://doi.org/10.1080/07908318.2020.1727495>
- Asker, B. (1998) Student reticence and oral testing: A Hong Kong study of willingness to communicate. *Communication Research Reports*, 15(2), 162-169. <https://doi.org/10.1080/08824099809362110>
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.
- Baghaei, P. (2013). Development and psychometric evaluation of a multidimensional scale of willingness to communicate in a foreign language. *European Journal of Psychology of Education*, 28(3), 1087-1103. <https://doi.org/10.1007/s10212-012-0157-y>
- Darasawang, P., & Reinders, H. (2021). Willingness to communicate and second language proficiency: A correlational study. *Education Sciences*, 11(9), 517. <https://doi.org/10.3390/educsci11090517>
- Burgoon, J. K. (1976). The unwillingness-to-communicate scale: Development and validation. *Communications Monographs*, 43(1), 60-69. <https://doi.org/10.1080/03637757609375916>

- Burgoon, J. K., & Burgoon, M. (1974). Unwillingness to communicate, anomia-alienation, and communication apprehension as predictors of small group communication. *The Journal of Psychology*, 88(1), 31-38. <https://doi.org/10.1080/00223980.1974.9915710>
- Donato, R. (1994). Collective scaffolding in a second language. In J. P. Lantolf, & G. Appel (Eds.), *Vygotskian approaches to second language research* (pp. 33-56). Ablex.
- Kang, S. J. (2005). Dynamic emergence of situational willingness to communicate in a second language. *System*, 33(2), 277-292. <https://doi.org/10.1016/j.system.2004.10.004>
- Lee, J. S., & Draijati, N. A. (2020). Willingness to communicate in digital and non-digital EFL contexts: Scale development and psychometric testing. *Computer Assisted Language Learning*, 33(7), 688-707. <https://doi.org/10.1080/09588221.2019.1588330>
- Lee, J. S., & Hsieh, J. C. (2019). Affective variables and willingness to communicate of EFL learners in in-class, out-of-class, and digital contexts. *System*, 82, 63-73. <https://doi.org/10.1016/j.system.2019.03.002>
- Lee, J. S., & Lee, K. (2020). Affective factors, virtual intercultural experiences, and L2 willingness to communicate in in-class, out-of-class, and digital settings. *Language Teaching Research*, 24(6), 813-833. <https://doi.org/10.1177/1362168819831408>
- Liu, M., & Jackson, J. (2008). An exploration of Chinese EFL learners' unwillingness to communicate and foreign language anxiety. *The Modern Language Journal*, 92(1), 71-86. <https://doi.org/10.1111/j.1540-4781.2008.00687.x>
- Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch Measurement Transactions*, 16(2), 878.

- Linacre, J. M. (2018). *Andrich thresholds: Disordered rating or partial credit structures (Categories or Andrich thresholds)*. Winsteps.
<https://www.winsteps.com/winman/disorder.htm>
- Long, M. H. (1981). Input, interaction, and second-language acquisition. *Annals of the New York Academy of Sciences*, 379(1), 259-278.
- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545-562. <https://doi.org/10.1111/j.1540-4781.1998.tb05543.x>
- MacIntyre, P. D., & Legatto, J. J. (2011). A dynamic system approach to willingness to communicate: Developing an idiodynamic method to capture rapidly changing affect. *Applied Linguistics*, 32(2), 149-171. <https://doi.org/10.1093/applin/amq037>
- McCroskey, J. C. (1992). Reliability and validity of the willingness to communicate scale. *Communication Quarterly*, 40(1), 16-25. <https://doi.org/10.1080/01463379209369817>
- McCroskey, J. C., & Baer, J. E. (1985, November 7-10). *Willingness to communicate: The construct and its measurement* [Conference presentation]. Annual Convention of the Speech Communication Association, Denver, CO, United States.
- Robson, G. G. (2015). The relationship between WTC and oral proficiency measurements in the study abroad context. *International Education Studies*, 8(12), 56-69.
<http://dx.doi.org/10.5539/ies.v8n12p56>
- Sato, R. (2020). Examining fluctuations in the WTC of [Language] EFL speakers: Language proficiency, affective and conditional factors. *Language Teaching Research*. Advance online publication. <https://doi.org/10.1177/1362168820977825>

- Surface, E. A., & Dierdorff, E. C. (2004). *Preliminary reliability and validity findings for the ACTFL writing proficiency test* (Report No. 2004-C04-R01). Surface, Ward, and Associates. <https://www.actfl.org/sites/default/files/assessments/acereports/ACTFL-WPT-Technical-Report-2004.pdf>
- Surface, E. A., & Dierdorff, E. C. (2003). Reliability and the ACTFL Oral Proficiency Interview: Reporting indices of interrater consistency and agreement for 19 languages. *Foreign Language Annals*, 36(4), 507-519.
- Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *Canadian Modern Language Review*, 50(1), 158-164. <https://doi.org/10.3138/cmlr.50.1.158>
- Thompson, G. L., Cox, T. L., & Knapp, N. (2016). Comparing the OPI and the OPIc: The effect of test method on oral proficiency scores and student preference. *Foreign Language Annals*, 49(1), 75-92. <https://doi.org/10.1111/flan.12178>
- Tschirner, E., & Bärenfänger, O. (2011). *An extension of inquiry into reliability issues of the ACTFL Writing Proficiency Test (WPT)* (Report No. 2011-US-PUB-1). Universität Leipzig. <https://www.actfl.org/sites/default/files/assessments/acereports/ACTFL-WPT-Technical-Report-2011.pdf>
- Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Appendix A: WTC Questionnaire

Each question is answered with a 100-point sliding scale. Participants must answer all questions from the first section and will be offered a random selection of three, four, and four questions from each of the following sections, respectively.

Willingness to Communicate in a Foreign-Language Scale (WTC-FLS)

The following questions are designed to measure your predisposition or readiness to communicate in [Language] when given the choice. You will answer questions based on **who** you would be communicating with and in **what** setting.

Section 1: Speaking [Language] with Native [Language] -Speakers In My Home Country

What percent of the time do the following statements describe you?

- 1.1 If I **encountered** some native speakers of [Language] in the street, restaurant, hotel, etc., I **hope an opportunity** would arise and they would talk to me.
- 1.2 If I **encountered** some native speakers of [Language] in the street, restaurant, hotel, etc., I **would find an excuse** and would talk to them.
- 1.3 If I **encountered** some native speakers of [Language] **who are facing problems** in my country because of not knowing our language, I **would take advantage** of this opportunity and would talk to them.
- 1.4 I **am willing** to accompany some native speakers of [Language] and be their tour guide for a day free of charge.
- 1.5 I **am willing** to talk with native speakers of [Language].
- 1.6 If someone **introduced me** to a native-speaker of [Language] I **would like to try** my abilities in communicating with him/her in [Language].

Section 2: Communicating in [Language] with Native [Language]-Speakers Online

What percent of the time do the following statements describe you?

- 2.1 If I **encountered** some native speakers of [Language] in an online forum, I **hope an opportunity** would arise and they would message me.
- 2.2 If I **encountered** some native speakers of [Language] in an online forum, I **would find an excuse** and would message them.
- 2.3 If I **encountered** some native speakers of [Language] **who are facing problems** online because of not knowing our language, I **would take advantage** of this opportunity and would message them.
- 2.4 I **am willing** to communicate with native speakers of [Language] online.
- 2.5 If someone **introduced me** to a native-speaker of [Language], I **would like to try** my abilities in communicating with him/her in [Language] online.

Section 3: Speaking [Language] with English-speakers in a School Setting In My Home Country

What percent of the time do (or would) the following statements describe you?

- 3.1 In order to practice my [Language], I **am willing** to talk in [Language] with my classmates **outside the class**.
- 3.2 I **am willing** to ask questions in [Language] **in the classes** at the university.
- 3.3 I **am willing** to talk and express my opinions in [Language] **in the class** when all my classmates are listening to me.
- 3.4 I **am willing** to have pair and group activities **in the class** so that I can talk in [Language] with my classmates.

- 3.5 In order to practice my [Language] I **am willing** to talk in [Language] with my professors **outside the class**.
- 3.6 I **am willing** to give a presentation in [Language] **in front of my classmates**.
- 3.7 In group work activities **in the class** when the group is composed of my **friends**, I **am willing** to speak in [Language].
- 3.8 In group work activities **in the class** when the group is **NOT composed of my friends**, I **am willing** to speak in [Language].

Section 4: Speaking [Language] with Native English-Speakers learning [Language] in the Mission Field

What percent of the time do (or would) the following statements describe you?

- 4.1 In order to practice my [Language], I **am willing** to speak in [Language] with native English-speaking missionaries learning [Language].
- 4.2 I **am willing** to ask questions in [Language] in **large group settings** of other native English-speaking missionaries learning [Language].
- 4.3 I **am willing** to talk and express my opinions in [Language] in **large group settings** of other native English-speaking missionaries learning [Language] that are **listening** to me.
- 4.4 I **am willing** to speak in [Language] in **pair and small group settings** with other native English-speaking missionaries learning [Language].
- 4.5 I **am willing** to teach a lesson in [Language] in front of a **large group** of other native English-speaking missionaries learning [Language].
- 4.6 I **am willing** to speak in [Language] with a group **composed of friends** that are also [Language] learners.

4.7 I **am willing** to speak in [Language] with a group **NOT composed of friends** that are also [Language] learners.