

Article

Not peer-reviewed version

A Generalized Technology Readiness Level Scale for Measuring Technology Maturity: Development and Pilot Validation Study

[Malcolm Townes](#) *

Posted Date: 29 November 2023

doi: 10.20944/preprints202311.1771.v1

Keywords: technology; technology maturity level; technology readiness level; technology transfer; gap fund; science and technology policy



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

A Generalized Technology Readiness Level Scale for Measuring Technology Maturity: Development and Pilot Validation Study [†]

Malcolm S. Townes

Washington University in St. Louis, St. Louis, Missouri, United States of America; malcolm.townes@att.net

[†] Parts of this paper were previously presented in the author's doctoral dissertation. This paper expands and extends that work.

Abstract: Recently considered modifications to public policy and trends in technology transfer practices have underscored the need for validated instruments to assess the maturity of technologies. This paper details a study that aimed to develop and present a generalized technology readiness level (GTRL) scale that a variety of professionals and researchers could use to assess the maturity of technologies in a range of contexts. Of particular interest is developing and validating an instrument that technology transfer professionals in federal laboratories, universities, and the private sector in the United States of America can use to assess the maturity of technologies to facilitate their technology transfer efforts. The primary goal of the research study was to highlight the need for validated measurement instruments for assessing the maturity of technologies in these contexts, test methods for validating the GTRL scale and other readiness level scales, and fill the knowledge gap regarding methods for properly validating such instruments. This included producing insights to inform future efforts to validate the GTRL scale as well as efforts to develop and validate other instruments to measure technology maturity. The study provides specific recommendations relevant to determining the face validity, content validity, inter-rater reliability, and intra-rater reliability of the GTRL scale and other such instruments for assessing technology maturity.

Keywords: technology; technology maturity level; technology readiness level; technology transfer; gap fund; science and technology policy

JEL Classification: D80; D81; L24; L29; L52; O14; O31; O32; O33; O38; O39

1. Introduction

The conveyance of technology created at federal laboratories and universities to the private sector for use that benefits the public good (i.e., technology transfer) has been an important public policy objective of the United States of America (U.S.) since the end of the Second World War. However, there is a measurement problem that is likely impeding progress. Technology, in this context, is broadly defined as culturally influenced information that social actors use to pursue the objectives of their motivations, and which is embodied in such a manner to enable, hinder, or otherwise controls its access and use (Townes, 2022).

The incidence of technology transfer has increased significantly since the passage of the Bayh-Dole Act of 1980, but progress seems to have stalled. Prior to the passage of the Bayh-Dole Act, only about five (5) percent of patented technologies derived from federally funded research were licensed (Schact, 2012). However, not all technologies are patentable, so it is reasonable to conclude that the percentage of technologies created that were licensed was likely lower. Currently, about 20 percent of technologies created at U.S. universities, many with the support of federal funding, are licensed (National Center for Science and Engineering Statistics, 2020; Townes, 2022; see Figure 1). The licensing percentage for technologies created at federal laboratories is about the same as that for

universities (National Institute of Standards and Technology, 2022; see Figure 2). However, it must be noted that the data are not based on a one-to-one match of technology disclosures with technology licenses and options. Technologies licensed and optioned in a given year are likely to have been disclosed in prior years.

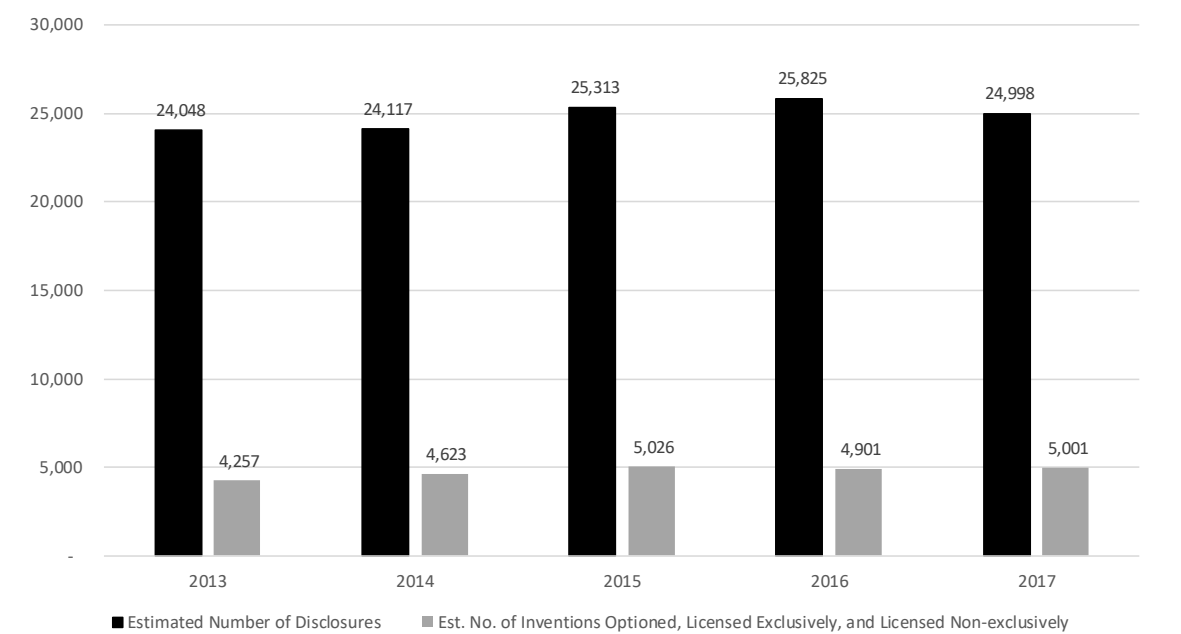


Figure 1. *University Technologies Transfer, 2013-2017.* Note. Figure created by author. Assumes an average of 1 unique technology per three 3 non-exclusive licenses. Data source: Hockstad, Mahurin, Miner, Porter, Robertson, & Savatski, 2017.

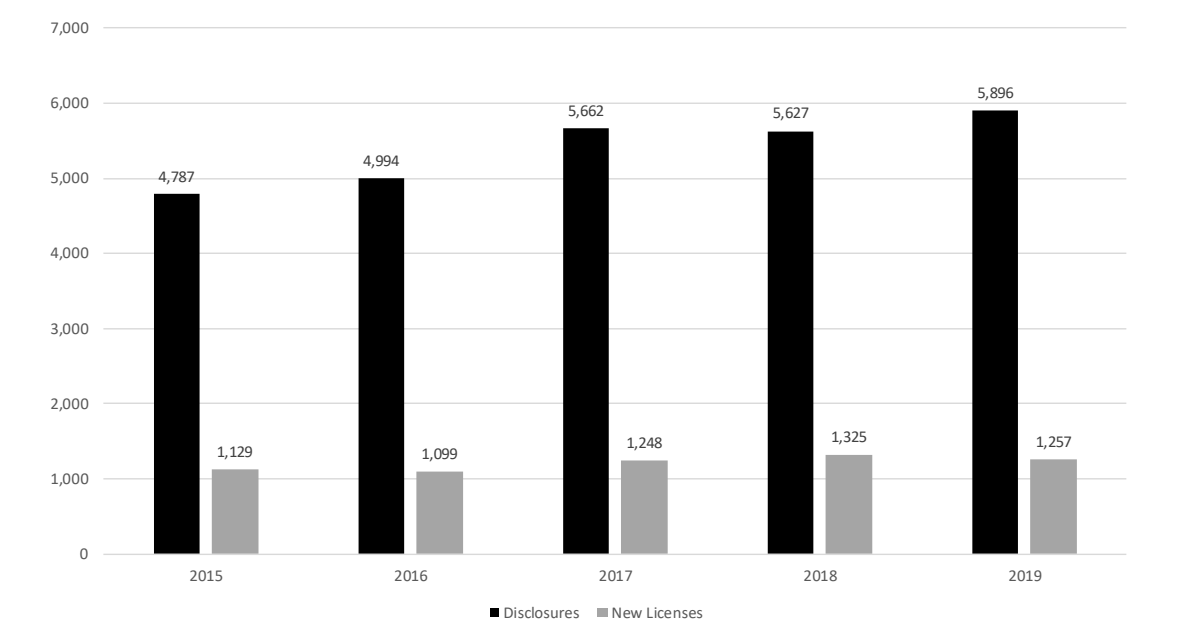


Figure 2. *Federal Laboratory Technology Transfer, 2015-2019.* Note. Figure created by author. Data source: National Institute of Standards and Technology, 2022.

Research suggests that technology maturity level is an important factor that influences the incidence of technology transfer (Juan, Wei, & Xiamei, 2010; Munteanu, 2012; Roussel, 1984; Townes,

2022,). University administrators and federal policymakers have either taken or considered actions related to technology maturity level to increase the incidence of technology transfer. Several universities have launched what are typically referred to as gap funds to de-risk and mature technologies to attract industry commercialization¹ partners (Munari, Rasmussen, Toschi, & Villani, 2016; Price & Sobocinski, 2002). The U.S. Congress has also considered intervention to increase the incidence of technology transfer. In the 117th Congress, Senate bill S.1260 and House bill H.R.2225 were introduced. Had they been enacted these bills would have authorized funding for universities to enable technology maturation.

The gap fund programs of U.S. universities and proposed public policy initiatives such as S.1260 and H.R.2225 underscore a measurement issue. How does one measure the maturity of a technology at the meso- and micro-level in a practical and useful way? Various approaches for measuring technology maturity have been proposed in the literature (see e.g., Lezama-Nicolás, Rodríguez-Salvador, Río-Belver, & Bildosola, 2018; Mankins, 1995; Mankins, 2009; Kriakidou, Michalakelis, & Sphicopoulos, 2013; Zorrilla, Ao, Terhorst, Cohen, Goldberg, & Pearlman, 2022). However, most of these approaches are intended for very specific contexts. Others are too complex and cumbersome to be of practical use to technology transfer practitioners. Moreover, none of these instruments, except for a few, have been validated in any meaningful way. Without a reliable measurement instrument, practitioners and policy makers cannot effectively determine whether or to what degree gap fund mechanisms and other interventions affect the incidence of technology transfer. It also makes it difficult, if not impossible, to synthesize various technology transfer research studies that have examined technology maturity as a predictor variable or to make reliable broad generalizations based on their results with any confidence.

Instruments and special apparatus capable of accurately and reliably measuring phenomena are essential for scientific advancement (see Kuhn, 2012). Being able to measure the maturity of a technology at a given point in time enables one to evaluate whether policies and programs are achieving their stated objectives of maturing technologies and whether achieving such objectives has an impact on increasing the incidence of technology transfer. As such, the aim of this study was to add to the knowledge base about approaches to properly validate instruments for measuring the maturity of technologies.

2. Review of the Related Literature

Perhaps the most well-known scale for characterizing the maturity level of a technology is the technology readiness level (TRL) scale developed by the National Aeronautics and Space Administration (NASA) in the United States of America. Some university technology transfer offices have begun applying this scale in some manner as part of their processes for vetting technologies and making portfolio decisions.

NASA employee Stand Sadin developed the original TRL scale in 1974 as a tool to help program managers estimate the maturity level of technologies being considered for inclusion in space programs (Banke, 2010; Mankins, 2009). In 1995, NASA employee John C. Mankins extended the scale from its original seven (7) ordinal levels to nine (9) and expanded the descriptions for each level (Mankins, 1995; Mankins, 2009). In 2001, after the General Accounting Office (GAO) recommended that the U.S. Department of Defense (DoD) make use of the scale in its programs, the Deputy Under Secretary of Science and Technology issued a memorandum endorsing the use of the TRL scale in new DoD programs. Guidance for using the scale to assess the maturity of technologies was incorporated into the Defense Acquisition Guidebook and detailed in the 2003 DoD Technology Readiness Assessment Deskbook (Nolte, 2008; Nolte & Kruse, 2011; Deputy Under Secretary of Defense for Science and Technology, 2003). In the mid-2000s, the European Space Agency (ESA)

¹ Although the terms *technology transfer* and *technology commercialization* are often used interchangeably, *technology commercialization* is used to specifically refer to technology transfer that is primarily profit driven (see Townes, 2022).

adopted a TRL scale that closely followed the NASA TRL scale (European Association of Research and Technology Organisations, 2014).

Technology transfer professionals have encountered challenges using the NASA TRL scale, which is not surprising. The NASA TRL scale is essentially a typology framework. Putting typologies into practice can be challenging (see Collier, Laporte, & Seawright, 2012). Also, as an agency of the federal government, NASA developed the TRL scale in the context of public sector applications. The public sector is not motivated by economic profit in the same way as the private sector. Moreover, technology development projects that have private sector applications are also likely to comprise a much broader range of types and kinds of technologies than those in any given federal agency in the public sector. Finally, there is no indication that anyone has ever established the validity and reliability of the NASA TRL scale. Consequently, developing and validating a generalized TRL scale would fill an important knowledge gap and could prove very useful in facilitating and advancing technology transfer research, policy, and practice.

The literature relevant to the topic of assessing technology maturity level is very limited. The construct of technology maturity level itself is difficult to define. Most of the reviewed literature about technology maturity level fails to explicitly define the construct (see e.g., Albert, 2016; Bhattacharya, Kumar, & Nishad, 2022; Chukhray, N., Shakhovska, N., Mrykhina, O., Bublyk, M., & Lisovska, L., 2019; Mankins, 1995; Mankins, 2009; Munteanu, 2012; Juan, Wei, & Xiamei, 2010). In a broad sense, one can think of technology maturity level as the phase of technological progress expressed as a function of some independent variable such as time, dollars spent refining the technology, number of workers utilizing the technology, or number of technical publications about the technology (O'Brien, 1962). Nolte (2008) resorted to analogies and scenarios to try to explain technology maturity level but never provided an exact definition. Nolte's concept of technology maturity level is inextricably tied to an instrumental definition of technology. Based on the discussion that Nolte offered, it seems reasonable to broadly define technology maturity level as the degree to which a technology is ready for use as intended to achieve an end that is acceptable.

It is worth noting that Nolte's conception of technology maturity level has the characteristics of value neutrality, context dependency, and multi-dimensionality (Nolte 2008). A given technology maturity level is neither "good" nor "bad" in and of itself. One's assessment of whether a given technology maturity level is acceptable depends entirely on the context in which one is using the technology (Nolte). A technology deemed to be at a given maturity level in one setting may be deemed to be at a different maturity level in another setting. Moreover, a comprehensive assessment of technology maturity level requires an examination from multiple perspectives (Nolte).

The literature suggests that the content domain of technology maturity extends beyond just technical development. The trajectory of technology is not just dictated by technical considerations (Stokes, 1997). A comprehensive assessment of technology maturity also needs to capture economic-related performance (Mankins, 2009). Market considerations also greatly influence the development and adoption of technology (Stokes). Adherents of the cultural school of technology studies would likely argue that culture too greatly influences the development and adoption of technology.

Technology maturity level seems closely associated with risk as the term is used in everyday language, which is more aptly called uncertainty². According to Blank and Dorf (2012), there are two primary types of uncertainty that commercial endeavors must manage. The first type is invention uncertainty (i.e., technical uncertainty), which is the possibility that the technology cannot be made to work as desired (Blank & Dorf). The second type is market uncertainty, which is the possibility that end users will not adopt the technology even if it can be made to work as desired (Blank & Dorf). Speser (2012) also discussed these differences in kinds of uncertainty and included firm-specific

² In an economic sense, uncertainty exists when the probability distribution of a future event is unknown or unknowable whereas risk exists when the probability distribution of a future event is measurable or known and thus the effects of the future event can be managed probabilistically (Knight, 1921).

uncertainty³ as a third type. Speser argued that one could not control market uncertainty but the lean startup methodology⁴, which has gained widespread acceptance among entrepreneurship practitioners and support organizations, calls this into question. Nolte (2008) argued that there are at least four dimensions of technology maturity level comprising technical, programmatic, developer, and customer viewpoints.

Given that most instances of technology transfer from federal laboratories and universities are intended to occur in the context of private sector commercial endeavors, it is logical to conclude that successful technology transfer also requires managing both technical and market uncertainty. Approaches that address market uncertainty without consideration of technical uncertainty will fail because they unduly raise hopes and make empty promises. They simply will not deliver. Those that address technical uncertainty without consideration of market uncertainty will fail in the market for lack of demand. No one will care. In both cases, the result is an unsuccessful attempt at technology transfer.

The NASA TRL scale is one of several approaches to operationalizing technology maturity level found in the literature. Use of the NASA TRL scale as a measure of maturity of a technology seems to be gaining traction in the field of technology transfer. Technology transfer professionals in the public sector use the NASA TRL scale (or an adaptation of it) because its use in NASA and DoD programs is Congressionally mandated. Many university technology transfer professionals also use the NASA TRL scale in some capacity to assess the maturity of technologies. This may be because the NASA TRL scale is the most well-known or it could be driven by the fact that the federal government funds much of the research conducted at U.S. universities who are likely taking their cues from the federal government.

The NASA TRL scale is not without its shortcomings. Olechowski, Eppinger, Tomascheck, and Joglekar (2020) investigated the challenges associated with using the NASA TRL scale in practice. Using an exploratory sequential mixed methods design consisting of qualitative semi-structured interviews and an online survey that included a best-worst scaling (BWS) experiment, they identified 15 challenges that practitioners face when using the NASA TRL scale. The participants in the study were predominantly private-sector practitioners from the aerospace, defense and government, and technology industries who had roles related to hardware development and advanced systems engineering (Olechowski, Eppinger, Tomascheck, & Joglekar). The study found that challenges encountered by practitioners were related to either system complexity, planning and review, or assessment validity. System complexity challenges pertained to incorporating new technologies into highly complex systems (Olechowski, Eppinger, Tomascheck, & Joglekar). This is likely the situation for most technology transfer endeavors in both the public and private sector. Challenges related to planning and review concerned the integration of NASA TRL assessment outputs with existing organizational processes, particularly those related to planning, review, and decision making (Olechowski, Eppinger, Tomascheck, & Joglekar). This seems to touch on the concept of absorptive capacity. Assessment validity challenges had to do with the reliability and repeatability of assessments using the NASA TRL scale (Olechowski, Eppinger, Tomascheck, & Joglekar). This suggests that the NASA TRL scale, as constituted, may be susceptible to idiosyncratic variation. This poses a potential impediment to advancing technology transfer theory and practice.

It is not surprising that private sector practitioners of technology transfer, even in the context of engagements with the federal government, would encounter challenges using the NASA TRL scale. As an agency of the federal government, NASA developed the TRL scale in the context of public sector applications. The public sector is not motivated by economic profit in the same way as the private sector. The NASA TRL scale focuses on technical uncertainty (i.e., invention uncertainty). As

³ Speser (2012) actually used the term *risk*.

⁴ The lean startup methodology is a philosophical and methodological approach to creating new ventures that was first articulated in its current form with the publication of *Four Steps to the Epiphany* in 2003 and the *Startup Owner's Manual* in 2012. Although it was primarily intended to apply to for-profit commercial enterprises it has since been applied to many types of endeavors (Blank, 2013).

such, it likely does not capture important economic and social factors relevant to technology development that are significant considerations for private sector decisions about opportunities to obtain and assimilate technologies from federal laboratories and universities.

Several scholars and practitioners have proposed alternative approaches to address shortcomings of the NASA TRL scale as well as alternate scales that express the notion of technology maturity level in various contexts or capture other aspects of its content domain (Table 1). Most of these scales also seem to focus on technical uncertainty. In fact, so many alternatives and variants of readiness level scales have been offered, introduced, or adapted for various situations that “readiness level proliferation” has become a problem in the public sector (Nolte & Kruse, 2011). However, few, if any, of these scales appear to have been validated in any scientifically meaningful way.

Table 1. Readiness Level Scales.

Accreditation Readiness Levels*	Learning Readiness Levels
Bilingualism Test Readiness Levels	Logistics Readiness Levels*
Business Readiness Levels	Love Readiness Levels
Change Readiness Levels	Manufacturing Readiness Levels*
Community Readiness Levels	Material Operational Readiness Levels
Condition of Readiness Levels	Operational Readiness Levels
Continuity Readiness Levels	Partner Readiness Levels
Countermeasures Readiness Levels*	People’s “Task” Readiness Levels
Defence Readiness Levels	Performance Readiness Levels
Defense Readiness Condition (DEFCON) Levels	Physical Readiness Levels
Demand Readiness Levels*	Primary Mental Abilities Readiness Levels
Direct Manufacturing Readiness Levels*	Problem Solving Readiness Levels
Disaster Readiness Levels	Programmatic Readiness Levels
Earthquake Readiness Levels	Reading Readiness Levels
Engineering & Manufacturing Readiness Levels*	Reuse Readiness Levels*
Entrepreneurship Readiness Levels	Risk Readiness Levels*
E-Procurement Readiness Levels	Security Readiness Levels
Extreme Heat Readiness Levels	Software Readiness Levels*
Fire Readiness Levels	Strategic Readiness Levels
Follower Readiness Levels	Supportability Readiness Levels*
Football Readiness Levels	Survival Readiness Levels
Fuel Readiness Levels	Sustainment Readiness Levels*
Funding Readiness Levels*	System Readiness Levels*
Human Effects Readiness Levels	Tactical Readiness Levels
Human Readiness Levels*	Technical Readiness Levels*
Innovation Readiness Levels*	Technology Readiness Levels*
Integration Readiness Levels*	TQM Readiness Levels
Internet Marketing Readiness Levels	Tropical Storm Readiness Levels
Investment Readiness Levels	University Technology Transfer Readiness Levels*
Investor Readiness Levels	Venture Readiness Levels

Note. Table created by author. Data sources: Fellnhöfer, 2015; Nolte & Kruse, 2011; Westerick, 2014. * indicates scale is related to technology maturity.

All of this suggests that there is a significant knowledge gap regarding the measurement and application of technology maturity in technology transfer policy and practice. To address this gap, I undertook an effort to develop a generalized technology readiness level (GTRL) scale that a variety of actors could use in a broad array of applications. This effort included conducting a pilot study of the validity and reliability of the GTRL scale.

Pilot studies are an important and valuable step in the process of empirical analysis, but their purpose is often misunderstood. In social science research, pilot studies can take the form of a trial run of a research protocol in preparation for a larger study or the pre-testing of a research instrument (Van Teijlingen & Hundley, 2010). The goal of a pilot study is to ascertain the feasibility of a methodology being considered or proposed for a larger scale study; it is not intended to test the hypotheses that will be evaluated in the larger scale study (Leon, Davis, & Kraemer, 2011; Van Teijlingen & Hundley). Unfortunately, pilot studies are often not reported because of publication bias that leads publishers to favor primary research over manuscripts on research methods, theory, and secondary analysis even though it is important to share lessons learned with respect to research methods to avoid duplication of effort by researchers (Van Teijlingen & Hundley).

The goal of the pilot study of the validity and reliability of the GTRL scale was to produce data and information to help answer several specific questions and take steps to fill the knowledge gap regarding approaches to properly validate instruments for assessing the maturity of technologies. These questions included: (1) What challenges are likely to be encountered in a larger study to assess the validity and reliability of the GTRL scale and other such instruments? (2) Can the methods for assessing content validity be applied to the GTRL scale? (3) Should participants in a larger validation study of the reliability of the GTRL scale be limited to university and federal laboratory technology transfer professionals? (4) How difficult will it be to recruit study participants? (5) How should a larger study familiarize participants with the GTRL scale? (6) What factors should be controlled for in a larger validity and reliability study? (7) How viable are asynchronous web-based methods for administering a validity and reliability study? (8) How viable is the approach of presenting marketing summaries of technologies to participants for them to rate? (9) How burdensome will it be for study participants to rate 10 technologies in multiple rounds? (10) How usable will the collected data be? and (11) What modifications, if any, to the GTRL scale should be considered before performing a larger validity and reliability study?

3. Data and Methods

The previous sections described the motivation for the study, situated the topic within the discourse on technology transfer, summarized what is known about technology maturity level and how it is measured, defined the knowledge gap, and stated the research questions that the study aimed to answer. This section describes the development of a generalized technology readiness level (GTRL) scale, the design and implementation of a pilot validity and reliability study, and testing of the intended procedures for analyzing the collected validity and reliability data.

3.1. Development of the Generalized TRL Scale

The NASA TRL scale was used as the starting point of reference for developing the GTRL scale because of its familiarity to technology transfer professionals and the simplicity of its application. The NASA TRL scale is an ordinal scale. Its application as an instrument for measuring technology maturity essentially treats technology maturity as a unidimensional construct. This single dimension is comprised of the concept (i.e., element) of “technology readiness”. The levels of the NASA TRL scale make up the items (i.e., indicators) that measure the “technology readiness” concept.

The content of various documents that discuss the NASA TRL scale and its application in other government agencies (see Assistant Secretary of Defense for Research and Engineering, 2011; Deputy Under Secretary of Defense for Science and Technology, 2003; Director, Research Directorate, 2009; Government Accountability Office, 2020) were reviewed to evaluate the comprehensiveness of the dimension indicators. Based on this content review, I added “Level 0” as an additional indicator to

capture the stage of ideation resulting in a 10-ordinal level scale (GTRL-10 scale). This affords the scale a ratio level of measurement quality that is likely to be useful for its application in research.

Defining the indicators and sub-indicators of each readiness level was an iterative process. It began with the evaluation of the definitions of each indicator to identify the gist of each. The objective of this step was to understand what should be included and excluded in the concept of each indicator. This was done by searching for higher order concepts (HOCs) and common themes that were adequate for characterizing each indicator regardless of the context. The HOCs and common themes identified were ideation, identification of relevant principles, thesis development, theoretical assessment, demonstration of feasibility, proof-of-concept, field testing, proof-of-outcomes, qualification testing, and ready for use. I specified generalized definitions in terms of these HOCs and common themes for each indicator of readiness on the 10-level ordinal scale. I then developed examples for each of three major contexts (medical drugs, medical devices, and non-medical applications) to provide additional guidance about the meaning of each indicator (i.e., scale level) and how it should be applied. As a final step in the process, I applied the indicators to various technologies perceived to be at different levels of maturity to check their fit. This process was repeated through several iterations until the definition of each indicator and context examples were deemed acceptable.

Once the indicators and context examples were defined, the focus shifted to the content and physical layout of the GTRL scale. To begin, I examined the layout of the original NASA TRL scale and its layout in various technology readiness assessment (TRA) deskbooks and guides to identify potential issues. In several of these resources, the TRL scale is presented on a single page. This limited space forces abbreviation of the definitions and requires the user to refer to other pages for additional information or examples for a given TRL level. Other layouts were multi-page presentations of the TRL scale where each page was dedicated to at most a few TRLs. However, the definitions of each level were still rather limited and abstract. The words on the pages were densely packed, which made skimming and rapid identification of specific information difficult. Examples were included to help illustrate the specific levels, but in some cases, they were dense and seemed to place a significant cognitive burden on the user. In other cases, the TRL scale was defined and illustrated in very narrow contexts.

The objective in designing the layout of the GTRL scale was to present it in a manner that was easy to follow and intuitive to minimize the cognitive burden on the user and the possibility of user error. To overcome the limitations identified in the various layouts and incarnations of the NASA TRL, I used Microsoft Office programs to format the GTRL with a multi-page layout that could be printed on standard letter size paper in landscape orientation (see Supplementary Information 1). Each indicator (i.e., readiness level) was printed in bold font on the left-hand side of the page. The definition of each level began immediately to the right of the readiness level. The definition was specified in terms of the HOC for the readiness level. Additionally, the indicator and definition were highlighted to make them easier to visually locate on the page.

Beneath each indicator and definition, a separate column dedicated to each of three major contexts was added to each level of the scale and examples were described for each context. The examples were slightly indented so that they began just to the right of the readiness level number, and no content would be immediately beneath the readiness level number to make locating each readiness level easier. The contexts were limited to medical drugs, medical devices, and non-medical to help minimize the cognitive burden for the user. Each context title was printed in bold font and surrounded by a border to make them easier to locate and read. Additionally, white space was used to visually separate each readiness level and to make the details of each readiness level easier to read.

With the format, layout, and content of the GTRL scale defined, attention turned to its validation. Assessing how accurately a scale or index measures a construct is important. Without establishing the validity and reliability of a scale or index, one cannot be sure if it is truly measuring what one hopes to measure or have confidence in the results of studies that use it. There are several types of validity that are used to establish the accuracy and reliability of an instrument for measuring a construct. Some are more relevant than others depending on the situation and context.

3.2. *Considering Face Validity*

In a general sense, validity is not a property of a measurement instrument but rather is an indication of the appropriateness for its use and interpretation in a given context (Knetka, Runyon, & Eddy, 2019). Researchers use different types of validity as evidence that they are in fact measuring the construct they claim to be measuring. One of the most basic types of validity is so-called face validity.

A review of the related literature suggests that assessing face validity of the GTRL scale is unnecessary. Face validity is an indication of whether the individuals taking the test or using the instrument consider it relevant and meaningful simply based on cursory review (Holden, 2010; Royal, 2016). An argument can be made that face validity is more fundamental than other types of validity such as construct validity and that the measures by which one establishes construct validity must have face validity to avoid the logically invalid epistemological arguments of regress and circularity (Turner, 1979). However, application of the philosophy of science concept of “correspondence rules” leads to the conclusion that neither face validity nor construct validity are required because constructing and testing new measures to evaluate a theoretical claim is an exercise in theory development and there is no need to make assertions about validity independent of theory (Turner). Moreover, it does not necessarily hold that an instrument or its indicators are erroneous in the absence of face validity (Holden). Even more, there is consensus among validity theorists that so-called face validity does not constitute scientific epistemological evidence of the accuracy of an instrument or its indicators (Royal).

3.3. *Assessing Content Validity*

Content validity is an indication of whether the construct elements (i.e., dimensions) and element items (i.e., indicators) of a measurement instrument are sufficiently comprehensive and representative of the operational definition of the construct it purports to measure (see Almasreh, Moles, & Chen, 2019; Bland & Altman, 2002; Yaghmaie, 2003). Content validity is considered a prerequisite for other types of validity (Almasreh, Moles, & Chen; Yaghmaie). For the GTRL scale, “technology readiness” is the only construct element. Each readiness level (i.e., TRL) is an indicator of the degree of “technology readiness” on the scale.

There is no universally agreed upon approach or statistical method for examining content validity (Almasreh, Moles, & Chen, 2019). The content validity index (CVI), proposed by Lawshe (1975), was chosen as the statistic for examining the content validity of the GTRL scale. Calculation of the CVI is a widely used approach to estimating content validity. Although there is the possibility that the CVI may overestimate content validity due to chance, it has the advantage of being simple to calculate as well as easy to understand and interpret.

Normal protocol for calculating the CVI for an instrument requires a panel of experts to rate the clarity and relevance of each element and indicator of the instrument (see Almasreh, Moles, & Chen, 2019; Lawshe, 1975). For purposes of evaluating the indicators, I replaced “relevance” with the concept of “usefulness” because theoretically an indicator can be relevant to a construct without necessarily being useful in a practical sense for measuring the concept. For evaluating the context examples for each indicator, I replaced “relevance” with the concept of “helpfulness” given the intended function of the examples.

The method for calculating the CVI for the GTRL scale consisted of recruiting several technology transfer professionals as domain experts to evaluate the clarity and usefulness of the construct element (i.e., technology readiness) and each element indicator (i.e., readiness level) and the clarity and helpfulness of each context example for an indicator. To recruit the domain experts, I sent a solicitation email message from my personal email account to qualified individuals who were familiar with me. No compensation was offered for participating in the study. Those who agreed to participate were sent a follow up email that contained a copy of the GTRL scale and a link to an online questionnaire (created on the Qualtrics platform) used to collect the content validation data (see Supplementary Information 2). The email instructed them to familiarize themselves with the GTRL scale before completing the questionnaire. Respondents did not have to complete the questionnaire

in one session. They could save their answers and return to complete the questionnaire later. However, if they did not continue or complete the questionnaire within two (2) weeks of the last time they worked on it, whatever answers they provided up to that point were recorded.

The overall process for assessing the content validity of the GTRL scale consisted of having an expert first rating the clarity of each indicator (i.e., readiness level) definition beginning with the lowest indicator followed by the next sequential indicator until the expert had rated every readiness level definition. Then the expert rated the clarity of each context example for each readiness level beginning with the lowest followed by the next sequential readiness level until the expert had rated all context examples for every readiness level. After rating clarity, the experts then rated the helpfulness of each context example as an aid to understanding how to assess whether a technology satisfied the requirements for a given readiness level beginning with the lowest followed by the next sequential readiness level until the expert had rated all context examples for every readiness level. The experts then rated the usefulness of each readiness level as a measure of the maturity of a technology, beginning with the lowest followed by the next sequential readiness level until the expert had rated every readiness level. Finally, the experts rated the usefulness of the concept of “technology readiness” as a measure of technology maturity.

This sequencing was adopted so that the experts would not have to keep changing their focus from one concept or feature to another. There was a concern that such focus shifting would affect how consistently the experts rated the scale features. Assessment of the element itself was done last because there was a concern that an expert’s response to this question would act to unduly prime the expert’s ratings of all other features of the scale.

The specifics of the content validity assessment began by showing the expert a matrix table with the definitions of the readiness levels listed in sequential order from lowest to highest and asking the expert to rate the clarity of the definition of each on the following Likert scale:

- 1 – not clear at all
- 2 – needs some clarification
- 3 – clearly understood

Next, the questionnaire presented the expert a matrix table with examples for each context for the lowest readiness level and instructed the expert to rate the clarity of each context example on the same Likert scale shown above. This process was repeated for each readiness level in sequential order until the experts had rated all context examples for all readiness levels.

After rating the clarity of each readiness level and its contexts example, the experts rated the helpfulness of each context example for each readiness level. Beginning with the lowest readiness level, the questionnaire presented the experts a matrix table with examples for each context and asked them to rate the helpfulness of each on the following Likert scale:

- 1 – not helpful at all
- 2 – marginally helpful
- 3 – helpful

This process was repeated for each readiness level in sequential order until the experts had rated the context examples of every readiness level.

The experts then rated the usefulness of each readiness level, as defined, as a measure of technology readiness. The questionnaire presented the experts a matrix table with the definitions of each readiness level listed in sequential order from lowest to highest and asked them to rate the usefulness of each readiness level on the following Likert scale:

- 1 – not useful at all
- 2 – marginally useful
- 3 – useful

Once the experts rated the clarity and usefulness of each readiness level and the clarity and helpfulness of each context example for each readiness level, the questionnaire asked them to rate the usefulness of the concept of “technology readiness” as a measure of technology maturity level using the same Likert scale used for rating the usefulness of the readiness levels above.

Finally, the questionnaire asked the experts if they were willing to be interviewed about their responses. It asked those who agreed to be interviewed to provide contact information.

3.4. Assessing Inter-Rater and Intra-Rater Reliability

The method for determining the inter-rater and intra-rater reliability of the GTRL scale relied on both domain experts and lay experts rating several technology summaries. Ten technologies at various stages of development were selected from public sources including the website of the Office of Technology Management (OTM) at Washington University in St. Louis (WUSTL), the SeedInvest platform (www.seedinvest.com), and the United States Patent and Trademark Office (USPTO). Using this information, a 1-to-2-page summary was created for each technology similar in style to technology summaries posted by many university technology transfer offices (see Supplementary Information 3). To ensure that a variety of maturity levels were represented, the information was modified in some cases to describe the technology as having achieved a greater or lesser degree of maturity.

A questionnaire was then created using Qualtrics to present the technology summaries and collect responses from raters (see Supplementary Information 4). The questionnaire consisted of an informed consent page, an instruction page, an overview of the GTRL scale layout with an option to print the GTRL scale for reference, and a page for each of the technology summaries with a response block presenting the readiness levels at the bottom.

To administer the survey, I solicited respondents via an email recruitment message sent using the Qualtrics platform to a total of 22 individuals who had experience assessing the maturity of technologies and were familiar with me personally. The career backgrounds of these individuals included technology transfer professionals, intellectual property attorneys, technical product managers, and technology entrepreneurs. No compensation was offered to them for participating in the study. The recruitment message informed the prospective raters that they would be rating 10 technologies and that at least two (2) weeks later they would be asked to rate the same 10 technologies again without referring to their previous ratings. A reminder message was sent five (5) calendar days after the original recruitment email message.

The prospective respondents were given 14 calendar days from the date of the initial recruitment email message to complete the questionnaire at which time the survey closed. Roughly 14 calendar days after the first survey closed, an email recruitment message for the second survey was sent via the Qualtrics platform to those individuals who responded to the first survey. A reminder message was sent seven (7) calendar days after the initial recruitment email message. The second questionnaire was essentially structured the same as the first questionnaire except that the respondents were allowed to skip the overview of the GTRL scale if they chose to do so (see Supplementary Information 5).

4. Results

The previous section described how the GTRL scale was developed and the methods for collecting data to assess the validity and reliability of the GTRL scale. This section briefly presents the results of piloting the data collection methodology described above as well as the planned methods for analyzing data to be collected in a larger study to assess the validity and reliability of the GTRL scale.

4.1. Content Validity

Two (2) individuals with experience in university technology transfer were recruited to serve as experts to assess the content validity of the GTRL scale. Each expert had more than 15 years of experience in university technology transfer and served as director of a university technology transfer office. Both experts completed the content validity questionnaire within one (1) week of being requested (see Table 2).

Table 2. Expert Ratings of Scale Level Definitions and Context Examples.

	Clarity		Useful/Helpful			Clarity		Useful/Helpful	
	Expert 1	Expert 2	Expert 1	Expert 2		Expert 1	Expert 2	Expert 1	Expert 2
TRL-0					TRL-5				
Definition	2	3	2	3	Definition	2	3	2	3
Med. Drugs	2	3	3	3	Med. Drugs	3	3	3	3
Med. Devices	3	3	3	3	Med. Devices	3	3	2	3
Non-Medical	3	2	3	3	Non-Medical	3	3	3	3
TRL-1					TRL-6				
Definition	3	3	3	3	Definition	3	3	3	3
Med. Drugs	3	2	3	2	Med. Drugs	2	3	3	3
Med. Devices	2	3	3	3	Med. Devices	3	3	3	3
Non-Medical	3	3	3	3	Non-Medical	3	2	3	2
TRL-2					TRL-7				
Definition	3	2	3	3	Definition	3	3	3	3
Med. Drugs	3	3	3	3	Med. Drugs	3	3	3	3
Med. Devices	3	3	3	3	Med. Devices	2	2	3	2
Non-Medical	2	3	3	3	Non-Medical	3	3	3	3
TRL-3					TRL-8				
Definition	3	3	3	3	Definition	3	3	3	3
Med. Drugs	3	2	3	2	Med. Drugs	3	3	3	3
Med. Devices	3	3	3	3	Med. Devices	2	2	2	2
Non-Medical	3	3	3	3	Non-Medical	3	3	3	3
TRL-4					TRL-9				
Definition	3	3	3	3	Definition	3	3	3	3
Med. Drugs	3	3	3	2	Med. Drugs	3	3	3	3
Med. Devices	3	3	3	3	Med. Devices	3	3	3	3
Non-Medical	2	3	3	3	Non-Medical	3	3	3	3

Note. Table created by author. 1 – not clear, useful, or helpful 2 – marginally clear, useful, or helpful 3 – clear, useful, or helpful.

The *CVratio* function in the *psychometric* package of the open-source programming language R in the RStudio developer environment was used to calculate the content validity ratios (CVRs) for each technology readiness level and context example (see Table 3). The scale level content validity index (CVI) was calculated as 0.8 using the CVRs for the usefulness of the technology readiness level definitions.

Table 3. Item Content Validity Ratios.

	Clarity	Helpful/Useful		Clarity	Helpful/Useful
TRL-0			TRL-5		
Definition	0.0	0.0	Definition	0.0	0.0
Med. Drugs	0.0	1.0	Med. Drugs	0.0	1.0
Med. Devices	1.0	1.0	Med. Devices	1.0	0.0
Non-Medical	0.0	1.0	Non-Medical	1.0	1.0
TRL-1			TRL-6		
Definition	1.0	1.0	Definition	1.0	1.0
Med. Drugs	0.0	0.0	Med. Drugs	0.0	1.0
Med. Devices	0.0	1.0	Med. Devices	1.0	1.0
Non-Medical	1.0	1.0	Non-Medical	0.0	0.0
TRL-2			TRL-7		
Definition	0.0	1.0	Definition	1.0	1.0
Med. Drugs	1.0	1.0	Med. Drugs	1.0	1.0
Med. Devices	1.0	1.0	Med. Devices	-1.0	0.0
Non-Medical	0.0	1.0	Non-Medical	1.0	1.0
TRL-3			TRL-8		

Definition	1.0	1.0	Definition	1.0	1.0
Med. Drugs	0.0	0.0	Med. Drugs	1.0	1.0
Med. Devices	0.0	1.0	Med. Devices	-1.0	-1.0
Non-Medical	1.0	1.0	Non-Medical	1.0	1.0
TRL-4			TRL-9		
Definition	1.0	1.0	Definition	1.0	1.0
Med. Drugs	1.0	0.0	Med. Drugs	1.0	1.0
Med. Devices	1.0	1.0	Med. Devices	1.0	1.0
Non-Medical	0.0	1.0	Non-Medical	1.0	1.0

Note. Table created by author.

4.2. Inter-Rater and Intra-Rater Reliability

The initial solicitation email message was distributed to a total of 22 individuals (see Table 4). Of the 22 individuals solicited, five (5) began the questionnaire and provided informed consent. However, only four (4) individuals rated one (1) or more technologies. Three (3) individuals rated all 10 technologies and one (1) person only rated four (4) of the technologies. Of the four (4) individuals who rated one (1) or more technologies, three (3) of them rated all 10 technologies a second time. Additionally, I rated all 10 technologies two (2) times with a delay period of at least two (2) weeks between each rating.

Table 4. Profile of Individuals Solicited for Inter-Rater and Intra-Rater Reliability Pilot Study.

Person	Sex	Technology Transfer Experience	Life Sciences Degree	Engineering, Comp. Sci., Mathematics, or Other Technical Degree	Social Sciences Degree	Technology Product Development Experience	Technology Entrepreneur	Technology Analyst or Other Technical Job
1**	Male	Yes	No	Yes	Yes	Yes	Yes	No
2	Male	No	No	Yes	No	No	No	Yes
3	Male	No	No	Yes	No	No	No	Yes
4	Male	No	No	Yes	No	No	No	Yes
5	Male	No	No	Yes	No	No	No	Yes
6	Female	Yes	Yes	No	No	No	No	No
7	Male	Yes	No	Yes	No	Yes	No	Yes
8	Male	No	No	Yes	No	No	No	Yes
9*	Male	No	No	No	No	No	Yes	Yes
10	Male	No	No	Yes	No	No	No	Yes
11	Male	No	No	Yes	No	No	No	Yes
12	Male	No	No	Yes	No	No	No	Yes
13*	Male	No	No	Yes	No	Yes	Yes	Yes
14	Male	No	No	Yes	No	Yes	Yes	Yes
15	Male	No	No	Yes	No	No	No	Yes
16	Male	Yes	No	No	Yes	No	No	No
17	Female	Yes	Yes	No	No	No	No	No
18	Male	Yes	No	Yes	No	No	No	Yes
19*	Male	No	No	No	No	No	No	Yes
20	Male	No	No	Yes	No	No	No	Yes
21*	Male	No	No	No	No	Yes	No	Yes
22	Female	Yes	No	No	No	No	No	No
23	Male	No	No	Yes	No	No	No	No

Note. * Indicates that the person participated in the study. ** Indicates the author.

Once the respondents submitted their ratings, the data were downloaded into Microsoft Excel spreadsheets, organized, and cleaned (see Table 5). The open-source programming language R in the RStudio developer environment was used to analyze the data (see Supplementary Information 6).

The *ICC* function for intraclass correlation coefficients (ICC) in the *psych* package as well as the *icc* and *kripp.alpha* functions in the *irr* package were used to calculate the inter-rater and intra-rater reliability of using the GTRL scale (see Tables 6–9). For the ICC calculations, the two-way mixed-effects model was selected because each technology was rated by the same set of raters. However, the raters were not randomly selected from a larger population of raters with similar characteristics. The single rater type was selected because the GTRL scale is intended to use the measurement from a single rater as the basis of assessment. Finally, the absolute agreement definition was chosen because the analysis focused on whether different raters assign the same score to the same subject. Additionally, the two-way mixed-effects single rater consistency ICC model was also included because it is identical to a weighted Cohen's kappa coefficient with quadratic weights for ordinal scales, which takes into consideration the magnitude of disagreement between raters (see Hallgren, 2012). However, Cohen's kappa can only be calculated for two raters whereas the ICC can be calculated for three or more raters.

Table 5. Rater Responses.

Technology	Round 1 GTRL Ratings					Round 2 GTRL Ratings				
	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
Tech 01	5	5	9	3	2	4	4	4	2	NR
Tech 02	9	6	3	6	4	8	6	3	3	NR
Tech 03	5	8	3	2	3	4	7	8	4	NR
Tech 04	4	3	4	9	3	4	4	1	6	NR
Tech 05	6	7	5	5	NR	6	8	6	7	NR
Tech 06	3	5	8	1	NR	3	3	3	1	NR
Tech 07	6	3	3	5	NR	5	6	6	5	NR
Tech 08	5	8	4	7	NR	6	8	2	7	NR
Tech 09	4	5	6	9	NR	3	5	5	9	NR
Tech 10	4	4	1	6	NR	4	4	3	6	NR

Note. NR indicates no response. Rater 1 is the author.

Table 6. Intraclass Correlation Coefficient for Inter-Rater Reliability Assessment.

R package	ICC	F	df1	df2	p	Lower	Upper
Round 1 (5 Raters; 4 Cases)							
irr ¹	-0.173	0.262	3	12	0.851	-0.232	0.356
irr ²	-0.2	0.262	3	4.62	0.85	-0.28	0.398
psych	0.0000	1.000	9	36	0.4577	-0.1360	0.3355
Round 1 (4 Raters; 10 Cases)							
irr ¹	-0.132	0.535	9	27	0.836	-0.249	0.187
irr ²	-0.137	0.535	9	23.3	0.834	-0.262	0.195
psych	0.0000	1.000	9	27	0.4635	-0.1833	0.3925
Round 2 (4 Raters; 10 Cases)							
irr ¹	0.209	2.06	9	27	0.0712	-0.058	0.615
irr ²	0.208	2.060	9	29.3	0.0678	-0.053	0.612
psych	0.2093	2.059	9	27	0.0712	-0.0575	0.6146

Note. * $p \leq 0.05$ ** $p \leq 0.01$ *** $p \leq 0.001$. ¹ Consistency model. ² Agreement model.

Table 7. Krippendorff's Alpha Reliability Coefficient for Inter-Rater Reliability Data.

	Round 1	Round 2
Statistic Value	-0.0538	0.2075
Number of Raters	5	4
Number of Cases	10	10
Number of Matches	44	120

Table 8. Intraclass Correlation Coefficient for Intra-Rater Reliability Assessment.

R package	ICC	F	df1	df2	P	Lower	Upper
irr ¹	0.551	3.45	39	39	0.0001***	0.292	0.734
irr ²	0.557	3.450	39	39	0.0001***	0.297	0.739
psych	0.5593	3.5385	39	39	0.0001***	0.3035	0.7399

Note. * $p \leq 0.05$ ** $p \leq 0.01$ *** $p \leq 0.001$. ¹ Consistency model. ² Agreement model.

Table 9. Krippendorff’s Alpha Reliability Coefficient for Intra-Rater Reliability Data.

Statistic Value	0.531
Number of Raters	2
Number of Cases	40
Number of Matches	80

5. Discussion

The preceding section presented the results of the data collection and analysis activities. This section aims to interpret those results. It answers the questions posed in the Introduction section, presents the major findings, and discusses conclusions drawn from those findings. It also addresses the limitations of the pilot study and recommendations for future studies of the validity and reliability of the GTRL scale and other instruments to measure technology maturity.

5.1. Findings

A key challenge with using the construct of technology readiness level as a measure of technology maturity is that there is no definition for a unit or quantity of “readiness.” In fact, the concept of “technology readiness” does not appear to be explicitly defined in the literature.

The pilot study demonstrated that the most popular method for assessing content validity can be successfully applied to assessing the GTRL scale and other readiness level scales. Although data from the pilot study are not sufficient to make broad generalizations, they do provide some insights. Depending on the reference, the scale level CVI value of 0.8 is considered acceptable for two expert raters. However, there may be a flaw in this epistemological approach to assessing content validity.

Use of the CVI to estimate content validity essentially assumes that experts are virtually omniscient and have near complete, infallible knowledge of the construct. This obviously cannot be true, and history is replete with examples. At one time, there were plenty of experts that considered the construct of race to be a factor of human intelligence, which has since been abandoned. Alternatively, prior to the early 2000s there were experts that had no idea that grit was a dimension of achievement and would not have missed its absence on an instrument to assess a person’s propensity for achievement and success. As such, when using expert-based approaches it is possible for an instrument to incorporate dimensions that experts deem critical, which are later demonstrated to be irrelevant. It is also possible for content validity to be deemed sufficiently high even if an instrument is missing relevant and useful dimensions of which the experts are simply not aware. At best, such a approaches can only estimate the degree to which the content of an instrument reflects the current consensus about dimensions of a phenomena, which could likely be determined just a readily with less time and effort through a thorough review of the literature.

The pilot study also demonstrated the viability of applying standard methods for assessing reliability to the GTRL scale and other readiness level scales. The ICC value for the inter-rater reliability of rating technologies had a p-value greater than 0.05 in both rounds. Thus, one could not reject the null hypothesis that there was no agreement among the raters. However, the ICC value for inter-rater reliability did increase from the first round of technology ratings to the second round. Krippendorff’s alpha reliability coefficient also increased from the first round of technology ratings to the second round, but was still well below the 0.667 threshold, which is considered the conceivable limit for tentative conclusions (see Krippendorff, 2004). The ICC value for estimating intra-rater reliability ranged from 0.557 to 0.559 and was statistically significant. This suggests that use of the scale is moderately stable over time. However, Krippendorff’s alpha reliability coefficient was 0.531,

which is less than the 0.667 threshold considered the smallest acceptable value. These results may indicate that modifications should be made to the instrument or the methodology, or both, before implementing a larger study of reliability as further discussed below.

It appears that the main challenge likely to be encountered in a larger study to assess the validity and reliability of the GTRL scale is obtaining the participation of technology transfer professionals. Six (6) of the 22 individuals (27%) solicited for the pilot study had experience as technology transfer professionals but none of them responded even though all were familiar with the researcher making the request. Technology transfer offices are notoriously understaffed and under-resourced. This may have contributed to the non-response and exacerbated societal trends that already tend to increase non-response. Of the 22 individuals solicited, several of them were hesitant to click on the link in the invitation email that was distributed via the online survey platform even though they recognized the name of the person making the request.

In a larger validation study of the reliability of the GTRL scale, it is probably advisable to control for the type of experience assessing technology maturity of the respondents. This can be accomplished by limiting participation in a given validation study to more homogenous groups, such as university technology transfer professionals. Although the GTRL scale is intended for a broad spectrum of users, validating the scale in more homogeneous groups will likely improve outcomes.

Although this was only a pilot study, one interpretation of the low inter-rater reliability statistics is that better instruction (in terms of type and amount) on the proper application of the scale will probably be needed in a larger study as well as when using the scale in various contexts. The nature and specifics of this instruction will likely need to vary a bit from group to group. For example, the type and amount of instruction required for university technology transfer professionals will likely differ to some non-trivial degree from that required for users in the private sector such as venture capital professionals. The approach taken in the pilot study probably will not suffice. A video providing a more detailed explanation of the GTRL scale with examples of its application is one option. Alternatively, one could also familiarize participants with the GTRL scale using a live synchronous overview that not only provides a more detailed explanation with examples of how to apply the scale but also enables participants to ask questions to clarify any confusion.

Asynchronous web-based methods for collecting data for studies of the validity and reliability of the GTRL scale and other instruments to measure technology maturity level appear to be viable. There were no issues administering the questionnaires using the web-based survey platform. Integrating a more detailed explanation of the GTRL scale using the approaches discussed above should not be a problem. Most web-based survey platforms can incorporate video media. Even if one chooses to use a live synchronous remote or in-person method to familiarize participants with the GTRL scale, it should not preclude the use of asynchronous web-based methods for data collection. Activities to familiarize participants with the GTRL scale can be unbundled from the technology rating and data collection activities.

Larger reliability studies of the GTRL scale should control for several factors. As discussed above, they should control the homogeneity of the technology assessment experience of the participants. Additionally, they should control for the prestige of the organization offering the technology and the researchers that created the technology. All these factors could potentially affect how participants rate a technology. Optionally, studies can also control for the category of technology. This may enable the collection of cleaner data for a more accurate assessment of inter-rater and intra-rater reliability for a given type of technology.

Presenting summaries of technologies to participants and having participants rate the technologies based on them is a viable approach for conducting reliability studies. The participants did not appear to encounter any problems. This approach has the advantage of mimicking the nature and structure of how demand-side technology transfer professionals obtain initial information about technologies. Moreover, it allows one to eliminate extraneous information and to control for various factors such as the prestige of the organization and the technology creators.

Having participants rate 10 technologies in multiple rounds did not appear to be overly burdensome. The longest time required to complete the questionnaire for the first round was within

four (4) days of beginning the questionnaire with most of the participants completing the questionnaire within two (2) days of beginning it. In the second round, all participants completed the questionnaire the same day they started.

The data that were collected employing the methodology of the pilot study was sufficiently usable. Very little cleaning of the data was necessary. Some data wrangling was necessary to get the data into the necessary format for specific functions of the R programming language.

There are possible modifications to the GTRL scale that it may be prudent to consider implementing before performing a larger validity and reliability study. First, it may be worthwhile to reduce the number of indicators (i.e., readiness levels) in the GTRL scale. This will likely increase the inter-rater and intra-rater reliability of the instrument. To illustrate this effect, consider a hypothetical scenario in which several people are asked to specify the outdoor conditions using the following 8-level ordinal scale rather than a thermometer:

- 1 – frigid
- 2 – cold
- 3 – chilly
- 4 – cool
- 5 – warm
- 6 – toasty
- 7 – hot
- 8 – scorching

Although the scale seems logical, it will likely be subject to significant idiosyncratic variation in its use resulting in very low inter-rater reliability and possibly low intra-rater reliability. The number of levels may simply be providing a false sense of precision. Moreover, it may be striving for a degree of accuracy that is not necessary for the indented use of the scale. Thus, the inter-rater reliability of the ordinal scale could likely be improved simply by reducing the number of indicators to the following 4-level ordinal scale:

- 1 – cold
- 2 – chilly
- 3 – warm
- 4 – hot

There would likely be less idiosyncratic variation with the more condensed scale because of greater agreement among individuals about the meaning of each indicator, a greater range of temperatures covered by each indicator, and fewer instances in which the temperature falls within the gray area between indicators.

The same scenario may be true of the GTRL scale and other readiness level measures. Assessments of technology maturity in the context of university and federal laboratory technology transfer may not need the level of precision and accuracy represented by the 10-level GTRL scale and other similar scales. The GTRL scale may prove more practical and useful for the intended contexts of its application if it is reduced from a 10-level ordinal scale to a 5-level ordinal scale (see Table 10 and Supplementary Information 7).

Table 10. Proposed 5-Level Ordinal GTRL Scale (GTRL-5).

Level	Maturation Milestone	Definition
0	Conceptualized	A perceived need or possible desired outcomes has been broadly described and a possible approach to satisfying the need or achieving the desired outcomes is conceived.

1	Theoretical feasibility demonstrated	The relevant phenomena and basic scientific principles have been identified. Experimental data and theoretical analysis demonstrate that it is feasible for the technology to achieve the desired outcome.
2	Proof-of-concept demonstrated	The elements of the technology have been integrated into prototypes comprising the key features, basic form, and desired functionality. Tests of the prototypes in laboratory and simulated environments provide additional evidence that the technology is likely able to achieve the desired outcome.
3	Application viability demonstrated	High-fidelity or full-scale prototypes close to the envisioned final form and functionality have been sufficiently tested in the intended environment or suitable facsimile environments and have achieved the desired outcomes with sufficient consistency.
4	Sufficiently mature for intended use	No further major modifications are necessary, and the technology is ready to be implemented in the intended environment.

It may also be worthwhile to reformat the examples for each readiness level and present them in a series of bullet points. This may help reduce the cognitive burden of using the instrument and allow users to more quickly locate the relevant information needed to rate a technology using the scale.

Finally, it may be helpful to include the higher order concept (HOC) or theme for each readiness level on the instrument. This should be done in a very conspicuous manner that makes it easy to visually locate. The HOC or theme may serve as a shorthand for users to help them mentally organize the details of the instrument and enable users to better apply the scale.

5.2. Limitations

As a pilot study, the primary objective of the study was to test key elements of an approach for assessing the validity and reliability of using a generalized technology readiness level (GTRL) scale as an instrument for evaluating the maturity of technologies, particularly in the context of university and federal laboratory technology transfer operations. The primary contribution of the study is mostly limited to generating insights about participant recruitment and the viability of the methodology. However, analysis of the data collected does provide some insights into the potential problems that may be encountered when using a 10-level GTRL scale for assessing the maturity of technologies even though the results of the data analysis are not suitable for making broad generalizations about using the GTRL scale.

5.3. Recommendations for Future Research

This investigation demonstrated a viable methodology for assessing the validity and reliability of a generalized technology readiness level (GTRL) scale as an instrument for evaluating the maturity of technologies in various contexts. On this basis, it is recommended that future research focus on larger scale studies of the validity and reliability of the instrument and other readiness level scales in specific contexts and with specific cohorts of participants with relatively homogenous technology assessment experience.

Researchers should consider whether expert-based approaches to assessing content validity are epistemologically reasonable. Those that employ the CVI method should increase the number of experts to between six (6) and nine (9) even though two (2) experts are considered minimally acceptable for such studies. Doing so should provide more confidence in the evaluation and minimize the probability of chance agreement. Additionally, future expert-based content validity studies should include an interview with willing experts about their responses as soon as possible after they complete the

questionnaire. Ideally, this follow-up interview should occur within two (2) weeks after the expert completes the questionnaire. It should focus on readiness levels and context examples the expert rated as 1 (not clear at all) and 2 (needs some clarification), context examples they rated as 1 (not helpful at all) and 2 (marginally helpful), and readiness levels they rated as 1 (not useful at all) and 2 (marginally useful). The interview should also ask the expert to expound on their rating of the usefulness of the concept of technology readiness as a measure of technology maturity regardless of their answer.

If future studies of inter-rater reliability present 10 technologies for rating, then the studies should have 10 participant raters to minimize the variance of the intraclass correlation coefficient (see Saito, Sozu, Hamada, & Yoshimura, 2006). Two rounds of ratings (i.e., replicates) should be sufficient for studies of intra-rater reliability. In such designs, between 10 and 30 participants should suffice (see Koo & Li, 2016). It is also recommended that future studies of validity and reliability offer financial compensation to incentivize the participation of technology transfer professionals and minimize non-response bias (see Dillman, Smyth, & Christian, 2014). Following up email recruitment messages with telephone calls may also be necessary to successfully recruit participants (see Dillman, Smyth, & Christian).

Finally, future efforts to develop easy to use, reliable instruments for assessing technology maturity should attempt to incorporate other dimensions of the construct. Ideally, these dimensions will be incorporated into a single instrument. For example, assessment of market uncertainty could be incorporated in TRL-2 and TRL-3 of the 5-ordinal level GTRL Scale (GTRL-5). Doing so may increase the predictive capability of the instrument while maintaining its ease of use and practicality.

6. Conclusion

This study aimed to develop and present a generalized technology readiness level (GTRL) scale that practitioners and scholars could use to assess the maturity of technologies in various contexts, particularly university and federal laboratory technology transfer to the private sector. It also aimed to test a methodology for assessing the validity and reliability of the GTRL scale and other readiness level instruments. The findings of this study demonstrate the viability of the methodology, highlight areas where the GTRL scale may be improved, and bring to light potential methodological issues that researchers may need to consider when conducting validity and reliability studies of the GTRL scale and other scales and instruments for assessing the maturity of technologies. This paper makes a valuable contribution to the technology transfer literature because it presents a practical scale for assessing technology maturity that has the potential to improve technology transfer research and practice. It demonstrates a robust replicable methodology for estimating the validity and reliability of readiness level scales for assessing the maturity of technologies in various contexts. This information can also help other researchers avoid “re-inventing the wheel” in future research efforts.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Funding: The author did not receive funding or any other form of support from any organization to assist with the submitted manuscript.

Data Availability Statement: The data generated or analyzed during this study are included in this article and its supplementary information files.

Acknowledgments: I would like to acknowledge the assistance of Mr. Keith D. Strassner and Mr. Eric W. Anderson, J.D., who served as domain experts for assessing the content validity of the generalized technology readiness level scale.

Competing Interests: The author has no relevant financial or non-financial competing interests to disclose.

References

1. Albert, T. (2016). *Measuring technology maturity: Operationalizing information from patents, scientific publications, and the web*. Springer Gabler.
2. Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, 15(2), 214-221. <https://doi.org/10.1016/j.sapharm.2018.03.066>

3. Assistant Secretary of Defense for Research and Engineering. (2011). *Technology readiness assessment (TRA) guidance*. United States Department of Defense. <https://apps.dtic.mil/sti/pdfs/ADA554900.pdf>
4. Bhattacharya, S., Kumar, V., & Nishad, N. (2022). Technology readiness level: An assessment of the usefulness of this scale for translational research. *Productivity*, 62(2), 112-124. <https://doi.org/10.32381/PROD.2021.62.02.2>
5. Bland, J. M., & Altman, D. G. (2002). Validating scales and indexes. *British Medical Journal*, 324(7337), 606-607. doi:<https://doi.org/10.1136/bmj.324.7337.606>
6. Blank, S. (2013). Why the lean start-up changes everything. *Harvard Business Review*, 91(5), 63-72. Retrieved from <http://www.coolprofs.com/FilePub/c49e3d77-47e0-4649-b155-ae4d0a7fdaa0/COOLUp+2014+-+Lean+startup+artikel.pdf>
7. Blank, S., & Dorf, B. (2012). *The startup owner's manual: The step-by-step guide for building a great company*. K&S Ranch, Inc.
8. Chukhray, N., Shakhovska, N., Mrykhina, O., Bublyk, M., & Lisovska, L. (2019). *Methodical approach to assessing the readiness level of technologies for the transfer*. Paper presented at the Conference on Computer Science and Information Technologies.
9. Collier, D., LaPorte, J., & Seawright, J. (2012). Putting typologies to work: Concept formation, measurement, and analytic rigor. *Political research quarterly*, 65(1), 217-232. <https://doi.org/10.1177/10659129124371>
10. Deputy Under Secretary of Defense for Science and Technology. (2003). *Technology readiness assessment (TRA) deskbook*. United States Department of Defense. <https://apps.dtic.mil/sti/pdfs/ADA418881.pdf>
11. Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. John Wiley & Sons.
12. Director, Research Directorate. (2009). *Technology readiness assessment (TRA) deskbook*. United States Department of Defense. <https://apps.dtic.mil/sti/pdfs/ADA524200.pdf>
13. European Association of Research and Technology Organisations [EARTO]. (2014). *The TRL scale as a research and innovation policy tool*. http://www.earto.eu/fileadmin/content/03_Publications/The_TRL_Scale_as_a_R_I_Policy_Tool_-_EARTO_Recommendations_-_Final.pdf
14. Fellnhöfer, K. (2015). Literature review: investment readiness level of small and medium sized companies. *International Journal of Managerial and Financial Accounting*, 7(3/4), 268-284. https://www.researchgate.net/profile/Katharina_Fellnhofer/publication/291670426_Literature_Review_Investment_Readiness_Level_of_Small_and_Medium_Sized_Companies/links/57e80d2908ae9e5e4558c179/Literature-Review-Investment-Readiness-Level-of-Small-and-Medium-Sized-Companies.pdf
15. Government Accountability Office. (2020). *Technology readiness assessment guide: Best practices for evaluating the readiness of technology for use in acquisition programs and projects* (GAO Publication No. GAO-20-48G). U.S. Government Printing Office. <https://www.gao.gov/assets/gao-20-48g.pdf>
16. Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23-34. <https://doi.org/10.20982/tqmp.08.1.p023>
17. Hockstad, D., Mahurin, R., Miner, J., Porter, K. W., Robertson, R., & Savatski, L. (2017). *AUTM 2017 licensing activity survey*. Association of University Technology Managers (AUTM). <http://www.autm.net>
18. Holden, R. R. (2010). Face validity. In *The Corsini encyclopedia of psychology*, (vol. 2, pp. 637-638). John Wiley & Sons. <https://doi.org/10.1002/9780470479216.corpsy0341>
19. Juan, Z., Wei, L., & Xiamei, P. (2010). *Research on technology transfer readiness level and its application in university technology innovation management*. Paper presented at the 2010 International Conference on E-Business and E-Government.
20. Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE—Life Sciences Education*, 18(1). <https://doi.org/10.1187/cbe.18-04-0064>
21. Knight, F. H. (1921). *Risk, uncertainty and profit*. New York, NY: Houghton Mifflin Company.
22. Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
23. Kuhn, T. S. (2012). *The structure of scientific revolutions* (4th ed.). The University of Chicago Press.
24. Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (second edition). Sage Publications.
25. Kyriakidou, V., Michalakelis, C., & Spicopoulos, T. (2013). Assessment of information and communications technology maturity level. *Telecommunications Policy*, 37(1), 48-62. <https://doi.org/10.1016/j.telpol.2012.08.001>
26. Lezama-Nicolás, R., Rodríguez-Salvador, M., Río-Belver, R., & Bildosola, I. (2018). A bibliometric method for assessing technological maturity: The case of additive manufacturing. *Scientometrics*, 117(3), 1425-1452. <https://doi.org/10.1007/s11192-018-2941-1>
27. Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel psychology*, 28(4), 563-575. Retrieved from https://www.researchgate.net/profile/Dr-Tulsi-Bhandari/post/Can_anybody_tell_me_the_difference_between_Content_Vaildity_Ratio_and_Content_Valid

- ity_Index/attachment/59d622bac49f478072e98fe4/AS%3A272118808285184%401441889587114/download/000+Lawshe_content_valdity.pdf
28. Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of psychiatric research*, 45(5), 626-629. <https://doi.org/10.1016/j.jpsychires.2010.10.008>
 29. Mankins, J. C. (1995). Technology readiness levels: A white paper (edited December 22, 2004). Washington, DC: Advanced Concepts Office, Office of Space Access and Technology, National Aeronautics and Space Administration (NASA). http://www.artemisinnovation.com/images/TRL_White_Paper_2004-Edited.pdf
 30. Mankins, J. C. (2009). Technology readiness assessments: A retrospective. *Acta Astronautica*, 65(9-10), 1216-1223. <http://www.onethesis.com/wp-content/uploads/2016/11/1-s2.0-S0094576509002008-main.pdf>
 31. Munari, F., Rasmussen, E., Toschi, L., & Villani, E. (2016). Determinants of the university technology transfer policy-mix: A cross-national analysis of gap-funding instruments. *The Journal of Technology Transfer*, 41, 1377-1405. <https://doi.org/10.1007/s10961-015-9448-1>
 32. Munteanu, R. (2012). Stage of development and licensing university inventions. *International Journal of Management and Enterprise Development* (1). <https://doi.org/10.1504/IJMED.2012.046796>
 33. National Center for Science and Engineering Statistics. (2020). *Survey of federal funds for research and development, fiscal years 2018-19*. National Science Foundation. Retrieved May 7, 2020 from <http://www.nsf.gov/statistics/fedfunds/>
 34. National Institute of Standards and Technology. (2022). *Federal laboratory technology transfer fiscal year 2019: Summary report to the President and the Congress*. U.S. Department of Commerce. <https://www.nist.gov/system/files/documents/2022/09/29/FY2019%20Federal%20Technology%20Transfer%20Report.pdf>
 35. Nolte, W. L. (2008). *Did I ever tell you about the whale?: Or measuring technology maturity*. Information Age Publishing.
 36. Nolte, W., & Kruse, R. (2011). *Readiness level proliferation*. Air Force Research Laboratory, Tech. Rep. 88ABW-2011-5501. https://ndiastorage.blob.core.usgovcloudapi.net/ndia/2011/system/13132_NolteWednesday.pdf
 37. O'Brien, M. P. (1962). Technological planning & misplanning, In J. R. Bright (Ed.), *Technological Planning on the Corporate Level* (pp. 73-99). Harvard University Press.
 38. Olechowski, A. L., Eppinger, S. D., Tomaszek, K., & Joglekar, N. (2020). Technology readiness levels: Shortcomings and improvement opportunities. *Systems Engineering*, 23(2). <https://doi.org/10.1002/sys.21533>
 39. Price, S. C., & Sobocinski, P. Z. (2002). Gap funding in the USA and Canada. *Industry and Higher Education*, 16(6), 387-392. doi:<https://doi.org/10.5367/0000000021012965>
 40. Roussel, P. A. (1984). Technological maturity proves a valid and important concept. *Research Management*, 27(1), 29-34. <https://doi.org/10.1080/00345334.1984.11756815>
 41. Royal, K. (2016). "Face validity" is not a legitimate type of validity evidence! *The American Journal of Surgery*, 212, 1026-1027.
 42. Saito, Y., Sozu, T., Hamada, C., & Yoshimura, I. (2006). Effective number of subjects and number of raters for inter-rater reliability studies, 25(9), 1547-1560. <https://doi.org/10.1002/sim.2294>
 43. Schacht, W. H. (2012). *The Bayh-Dole act: selected issues in patent policy and the commercialization of technology*. (RL30276). Library of Congress. <http://crsreports.congress.gov>
 44. Speser, P. L. (2012). *The art and science of technology transfer*. John Wiley & Sons.
 45. Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Brookings Institution Press.
 46. Townes, M. S. (2022). *The influence of technology maturity level on the incidence of university technology transfer and the implications for public policy and practice* (Publication No. 29209868) [Doctoral dissertation, Saint Louis University]. ProQuest.
 47. Turner, S. P. (1979). The concept of face validity. *Quality and Quantity*, 13(1), 85-90. <https://link.springer.com/content/pdf/10.1007/BF00222826.pdf>
 48. Van Teijlingen, E., & Hundley, V. (2010). The importance of pilot studies. *Social Research Update*, 35(4), 49-59. <https://sru.soc.surrey.ac.uk/SRU35.html>
 49. Westerik, F. H. (2014). *Investor readiness: Increasing the measurability of investor readiness* (Master's thesis, University of Twente). https://essay.utwente.nl/64605/1/Westerik_MA_MB.pdf
 50. Yaghmaie, F. (2003). Content validity and its estimation. *Journal of medical education*, 3(1), 25-27. <https://brieflands.com/articles/jme-105015.pdf>
 51. Zorrilla, M., Ao, J. N., Terhorst, L., Cohen, S. K., Goldberg, M., & Pearlman, J. (2022). Using the lens of assistive technology to develop a technology translation readiness assessment tool (TTRAT)TM to evaluate market readiness. *Disability and Rehabilitation-Assistive Technology*, 16. <https://doi.org/10.1080/17483107.2022.2153936>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.