



## Validation of a trust in medical technology instrument

Enid Montague\*

Industrial and Systems Engineering Department, University of Wisconsin-Madison, 3217 Mechanical Engineering, 1513 University Avenue, Madison, WI 53706, USA

### ARTICLE INFO

#### Article history:

Received 12 January 2009

Accepted 23 January 2010

#### Keywords:

Trust  
Health care  
Technology

### ABSTRACT

A patient's trusting attitude towards technology used in their medical care may be a predictor of acceptance or rejection of the technology and, by extension, the physician. The aim of this study was to rigorously determine the validity of an instrument for measuring patients' trust in medical technology. Instrument validity was established based on a framework, which included test and data evidence for validity assessment. The framework for validity assessment evaluates the instrument on content, substantive, structural, generalizability, external and consequential aspects of validity. The results of the current study show that the instrument is reliable and valid for assessing a patient's trust in obstetric medical technology.

© 2010 Elsevier Ltd. All rights reserved.

### 1. Introduction

In the medical domain, understanding the work system effects of patient–physician interpersonal trust relationships (Pearson and Raeke, 2000) and how to design medical technologies for lay users (Martin et al., 2008) have been identified as important research questions. Patient safety researchers have also advocated for a socio-technical systems approach to solving health system problems, which include patients as members of the works system (Buckle et al., 2006). Developing tools to collect empirical data about patients' perceptions of their health care system is the first step towards understanding the relationship between patients' perceptions and the functioning of the work system. Valid measurement tools are important as we begin to think of how to design better, safer and more efficient work systems that include multiple types of users (patients and care providers). This study describes the validation of a tool to measure patients' trust in medical technology, with the future goal that this tool will be used to characterize patient–physician relationships and patients' request, acceptance, and rejection of technologies used in their care.

The development of the trust in medical technology instrument involved defining variables, modeling observations, and evaluating the measures. Work with variables involved defining constructs within a theoretical framework, identifying indicators of the construct, and mapping those indicators onto the theoretical framework. Modeling the observations involved specifying rules for converting observations into numbers, creating a mathematical model of how those numbers could be combined to create

measures, and specifying a frame of reference for interpreting measures. The evaluation of measures was completed by reliability and validity analyses that ensured measures were stable across multiple contexts and consistent with the framework within which they were created.

#### 1.1. Trust

Trust, a person's level of belief in a person or thing, is a fundamental attribute in all relationships. A variety of trust relationships exist; trust between two or more people (interpersonal trust) (Larzelere and Huston, 1980), a person's trust with a system or institution (social trust) (Castelfranchi and Falcone, 2002) and a person's trust with a technology or device (technological trust) (Muir, 1994). Humans use metrics of trust to determine which humans they form relationships with, which institutions they want to be a part of and which technologies they use.

Interpersonal trust has been operationalized as comprised of the factors predictability, reliability and dependability (Larzelere and Huston, 1980), as stages of predictability, dependability and faith (Rempel et al., 1985), and as a behavioral model (DeFuria, 1996). DeFuria (1996) developed the interpersonal trust surveys from a behavioral model of interpersonal trust, which defines the behaviors that lead to increased or decreased trust between individuals (see Table 1). DeFuria (1996) argues that trust situations always involve vulnerability, risk, and expectations of the other person's trustworthy motivation and competence. The concept of interpersonal trust is explored across domains though differences exist in how the construct is measured and defined in work systems such as education (Tschannen-Moran and Hoy, 2000) and health care (Pearson and Raeke, 2000).

\* Tel.: +1 608 890 2562; fax: +1 608 262 8454.

E-mail address: [emontague@wisc.edu](mailto:emontague@wisc.edu)

**Table 1**  
Trust enhancing and reducing behaviors.

Trust Enhancing Behaviors	Trust Reducing Behaviors
Sharing relevant information	Distorting, withholding, or concealing real motives
Reducing controls	Falsifying relevant information
Allowing for mutual influence	Attempting to control or dominate
Clarifying mutual expectations	Obscuring, distorting, or avoiding the discussion of mutual expectations
Meeting expectations	Not meeting the trusting individuals expectations of performance or behavior
	Attempting to evade responsibility for behavior

Table from DeFuria (1996). Interpersonal trust surveys: Jossey-Bass.

### 1.2. Patient trust in physician

Patient trust in physician is a multidimensional construct that has different definitions between and within disciplines. Thom et al. (2002) define patient trust in physician as a set of beliefs or expectations that a care provider will perform in a certain way, while Pearson and Raeke (2000) define trust as an emotional characteristic, where patients have a comforting feeling of faith or dependence in a care provider's intentions. Several scales have been developed to measure patients' trust in their care provider, but none have explored the construct as a sociotechnical construct (Kao et al., 1998; Thom et al., 1999, 2002; Pearson and Raeke, 2000; Zheng et al., 2002; Boehm, 2003). Pearson and Raeke's (2000) review paper found the factors; competence, compassion, privacy, confidentiality, reliability, dependability, and communication to be the most common factors of patient trust in physician. Satisfaction is sometimes conflated with trust as it is also considered an important measure of quality in health systems, however, Thom et al. (1999) found patient trust in physician to be distinct from patient satisfaction. Patient trust in physician has been explored in relation to patient outcomes and behaviors such as adhering to medical advice, malpractice litigation and seeking healthcare services. While patient trust in physician has been found to be an important predictor of patient health and quality outcomes some researchers hypothesize that changes in healthcare services and practices may be undermining the trust relationship between patients and physicians (Pearson and Raeke, 2000). Other researchers argue that enhancing patient trust in physicians may be a solution to existing problems in the provision of health care (Boehm, 2003). In healthcare systems the concept of interpersonal trust cannot be separated from organizational, social, and technological trust (Montague, 2009). In order to understand how changes in health care service (such as technology implementation) might affect trust and how patient trust in physician may be an effective solution for health system problems a theoretical understanding of trust in health systems is needed. Therefore measures of trust are needed at the interpersonal, social and technological level.

### 1.3. Trust in technology

Researchers in consumer studies, psychology, engineering, and information systems have looked at trust relationships between users and technologies. Trust and technology scholarship has largely focused on trust between users and automation technologies (Lee and Moray, 1994; Muir, 1994; Muir and Moray, 1996; Parasuraman and Riley, 1997; Bisantz and Seong, 2001; Hoc, 2000; Lewandowsky et al., 2000; Moray, 2000; Sheridan, 2002; Dzindolet et al., 2003; Lee and See, 2004) and trust between users and internet technologies (Basso et al., 2001; Corbitt et al.,

**Table 2**  
Eigenvalues of all components.

Factor	Eigenvalue	Percent	Cum Percent
1	31.1732	38.967	38.967
2	12.2743	15.343	54.309
3	5.5092	6.886	61.196
4	2.6888	3.361	64.557
5	2.4134	3.017	67.574
6	2.1051	2.631	70.205
7	1.8676	2.334	72.539
8	1.6205	2.026	74.565
9	1.4563	1.82	76.385
10	1.3461	1.683	78.068
11	1.1506	1.438	79.506
12	1.0942	1.368	80.874
13	1.0269	1.284	82.158
14	0.9483	1.185	83.343
15	0.8488	1.061	84.404

2003; Grabner-Krauter and Kaluscha, 2003; McKnight et al., 2003; Kong and Hung, 2006; McKnight and Choudhury, 2006; Awad and Ragowsky, 2008). This paper explores patients' trust in medical technology and the validation of an instrument to measure patients' trust in medical technology.

### 1.4. Trust in medical technology

Patient trust in medical technology may be an important factor of functional work systems; particularly as health care work systems move to higher reliance on and use of medical technologies. Understanding patients' trust in medical technologies may also provide insight into optimized patient–physician relationships. Trusting attitudes towards in medical technologies has been explored from both provider and patient perspectives. Moffa and Stokes (1997) explored health worker trust in medical expert systems, to gain insight into the importance of work system domain

**Table 3**  
Factor one eigenvalues, item number and text.

Loading	Item #	Item text
0.812036	Q1	The technology was accurate.
0.883677	Q2	The technology was trustworthy.
0.876487	Q3	The technology was reliable.
0.669783	Q4	The technology was safe.
0.785337	Q5	The technology had reliability.
0.828666	Q6	The technology was precise.
0.790656	Q8	The technology was honest.
0.820144	Q9	I trust the technology.
0.867855	Q10	I had positive feeling about the technology.
0.819439	Q11	The technology was responsible.
0.647163	Q12	The technology behaved responsibly.
0.86061	Q13	The technology was successful.
0.82544	Q14	The technology performed accurately.
0.834301	Q15	I felt secure because of the technology.
0.792787	Q16	The technology was secure.
0.418035	Q17	I was well informed about the technology.
0.74693	Q18	The technology was well informed.
0.821281	Q19	The technology gave real results.
0.849689	Q20	I felt confident that the technology was working.
0.809519	Q21	The technology gave correct results.
0.504104	Q23	I believe the technology was well researched.
0.487355	Q24	The technology was well researched.
0.836129	Q25	I had confidence in the technology
0.809389	Q26	The technology was helpful.
0.80656	Q28	Using the technology gave me security.
0.850788	Q29	The technology was effective.
0.638151	Q30	The technology made/kept me healthy.
0.580229	Q31	I relied on the technology.

**Table 4**  
Factor two eigenvalues, item number and text.

Loading	Item #	Item text
0.640448	Q32	My health care provider was accurate.
0.785753	Q33	My health care provider was trustworthy.
0.862042	Q34	My health care provider was reliable.
0.632924	Q35	My health care provider behaved in a safe manner.
0.626033	Q36	My health care provider was precise.
0.37382	Q37	I knew a lot about my health care provider.
0.820321	Q38	My health care provider was honest.
0.817634	Q39	I trust my health care provider.
0.874118	Q40	I feel good about the health care provider I had.
0.837499	Q41	I have positive feelings about my health care provider.
0.824046	Q42	My health care provider was honest.
0.807036	Q43	My health care provider was responsible.
0.801116	Q44	My health care provider behaved responsibly.
0.359269	Q45	My health care provider was successful with regards to my treatment.
0.728892	Q46	My health care provider was well informed.
0.574001	Q47	My health care provider gave real results.
0.841816	Q48	I had confidence in my health care provider.
0.618576	Q49	My health care provider gave me the correct information about my health situation.
0.258403	Q50	I researched my health care provider.
0.74181	Q51	I had confidence in my health care provider.
0.750931	Q52	My health care provider was helpful.
0.777136	Q53	My health care provider had integrity.
0.570084	Q54	My health care provider cared about me.
0.124698	Q55	My health care provider affected my life.
0.694556	Q56	My health care provider was effective.
0.324146	Q57	My health care provider made or kept me healthy.

in user formation of trust in technologies. Kjerulff et al. (1992) also looked at health workers' trust in medical technologies in their study of nurses' technology anxiety or fear of working with medical technology. They found trust in medical technologies to be related to the department nurses worked in, and attitudes about their work, such as satisfaction, stress and interpersonal relationships.

Trusting attitudes towards medical technologies has also been researched from the patient's perspective. Timmons et al. (2008) conducted a study of trust in medical technologies, where their primary research questions was 'how do lay users come to trust automatic external defibrillators'. They found that when confronted with using an unfamiliar medical technology; users' trust in the technology is constructed through a combination of trust in the technology, people and institutions. Montague et al. (2009) developed an empirical model of the construct trust in medical technology for patients. They found trust in technologies in medical domains to be distinct from trust in technologies in other domains.

Trust in medical technology has also been explored as a social measure. Calnan et al. (2005) conducted a national survey to assess public attitudes about a variety of innovative health care technologies. They found general public ambivalence about new medical technologies. However, respondents who reported consistent negative responses regarding new medical technologies were more likely to report distrusting attitudes about science, health care and care providers. Calnan et al.'s (2005) findings provide some evidence for a sociotechnical systems understanding of the construct. Individual differences in attitudes about medical technologies have been explored as well; Groeneveld et al. (2006) found that people of differing ethnic groups may have differing attitudes about using innovative medical technologies. Specifically, they found that Whites reported higher degrees of acceptance of medical technology innovations than Blacks. Similar sociocultural differences have been reported in patient trust in care provider and institution, which provides support for beliefs that patient trust in medical technology may be related to other types of patient trust (Rose et al. 2004).

**Table 5**  
Factor three eigenvalues, item number and text.

Loading	Item #	Item text
0.770999	Q58	My health care provider used the medical technology accurately.
0.763641	Q59	My health care provider used the technology in a trustworthy manner.
0.608639	Q60	My health care provider made sure the technology was reliable.
0.835192	Q61	My health care provider used the technology in a safe way.
0.806665	Q62	My health care provider was knowledgeable about the technology.
0.78116	Q63	I trust how my health care provider used the technology.
0.74261	Q64	I feel good about the way my health care provider used the technology.
0.664132	Q65	I have positive feelings about how my health care provider used the technology.
0.89806	Q66	My health care provider used the technology in a responsible manner.
0.853382	Q67	My health care provider used the technology responsibly.
0.767558	Q68	My health care provider was successful in their use of the technology.
0.859124	Q69	My health care provider used the technology accurately.
0.770122	Q70	My health care provider was well informed about the medical technology.
0.539829	Q71	My health care provider used the technology to give real results.
0.880509	Q72	My health care provider was confident using the medical technology.
0.526948	Q73	The results my health care provider obtained from the technology were correct results.
0.425647	Q74	My health care provider used the medical technology to give correct results.
0.817842	Q75	My health care provider used the technology with confidence.
0.173468	Q76	My health care provider believed the technology was helpful.
0.784565	Q77	My health care provider used the technology in a caring manner.
0.670522	Q78	My health care provider used the technology effectively.
0.236036	Q79	My health care provider used the technology to make me or keep me healthy.
-0.077502	Q80	My health care provider had high reliance on the medical technology.

### 1.5. Patient trust

Studies of patient interpersonal relationships have alluded to a missing dimension of the construct patient trust in care provider. An observational study examined patients' perceptions of their physician's interpersonal manner through ratings such as satisfaction, trust, and knowledge of patient and autonomy support (Franks et al., 2005). The researchers found a relationship between patients' perceptions of their physician and health status decline. Through multilevel analyses, the researchers believe that this relationship is not a physician effect, but may be the result of another confounding variable (Franks et al., 2005). Patient's trust in technology may be a variable that provides additional insight into the factors that predict patient trust or distrust in their physician. Other studies have looked at patients' assessments of their care provider and linked these ratings with changes in health status, patient satisfaction, and adherence to advice (Franks et al., 2005). Franks et al. (2005) argue that patients are likely to report being more satisfied with their care provider if they have better health status and conclude these results may not be related to care provider traits, indicating a need for more dynamic measures of physician traits. While Franks et al. (2005) recommend assessment of psychological and personality traits to understand patient trust in care provider, attitudes towards technology would be an insightful addition.

In a study of patient trust in care providers Fung et al. (2005) found that patients report a preference for physicians with

**Table 6**  
Items flagged for further evaluation.

#	Target	Item	rpm	M <sub>Sw</sub>	Z <sub>w</sub>	M <sub>su</sub>	Z <sub>u</sub>	Decision
1	Tech	The technology was accurate.	<b>0.05</b>	<b>2.82</b>	<b>5.30</b>	<b>5.57</b>	<b>7.60</b>	Drop/revise
3	Tech	The technology was reliable.	0.80	0.50	–3.30	<b>0.37</b>	–2.60	Drop/revise
8	Tech	The technology was honest.	<b>0.29</b>	<b>2.10</b>	<b>4.00</b>	<b>2.94</b>	<b>2.90</b>	Drop
11	Tech	The technology was responsible.	0.80	0.66	–2.20	<b>0.48</b>	–2.30	Drop
12	Tech	The technology behaved responsibly.	0.80	0.59	–2.80	0.53	–2.10	Keep
14	Tech	The technology performed accurately.	0.80	<b>0.46</b>	–3.50	<b>0.34</b>	–2.50	Drop/revise
18	Tech	The technology was well informed.	<b>0.45</b>	<b>1.78</b>	<b>3.30</b>	<b>1.99</b>	<b>2.20</b>	Drop/revise
21	Tech	The technology gave correct results.	0.81	0.53	–3.10	<b>0.38</b>	–2.60	Drop/revise
23	Tech	I believe the technology was well researched.	<b>0.15</b>	<b>3.62</b>	<b>9.90</b>	<b>4.31</b>	<b>9.40</b>	Drop/revise
26	Tech	The technology was helpful.	0.79	0.59	–2.60	<b>0.45</b>	–2.10	Drop/revise
28	Tech	Using the technology gave me security.	<b>0.42</b>	<b>2.09</b>	<b>5.40</b>	<b>3.05</b>	<b>6.80</b>	Drop/revise
30	Tech	The technology made/kept me healthy.	0.82	<b>0.42</b>	–3.90	<b>0.32</b>	–2.70	Drop/revise
32	Prov	My health care provider was accurate.	0.54	<b>1.81</b>	<b>4.30</b>	<b>3.52</b>	<b>6.70</b>	Drop/revise
34	Prov	My health care provider was reliable.	0.58	0.56	–1.40	<b>0.25</b>	–1.10	Keep
35	Prov	My health care provider behaved in a safe manner.	0.63	0.66	–1.10	<b>0.19</b>	–1.70	Keep
36	Prov	My health care provider was precise.	0.57	0.77	–0.60	<b>0.23</b>	–1.10	Keep
38	Prov	My health care provider was honest.	0.51	<b>1.85</b>	<b>3.30</b>	<b>1.89</b>	<b>2.50</b>	Drop/revise
39	Prov	I trust my health care provider.	0.58	<b>0.46</b>	–1.70	<b>0.11</b>	–1.40	Keep
40	Prov	I feel good about the health care provider I had.	0.69	<b>0.46</b>	–2.20	<b>0.18</b>	–2.10	Drop/revise
41	Prov	I have positive feelings about my health care provider.	0.67	<b>0.47</b>	–2.10	<b>0.18</b>	–1.90	Drop/revise
42	Prov	My health care provider was honest.	0.66	0.62	–1.40	<b>0.23</b>	–1.80	Keep
45	Prov	My health care provider was successful with regards to my treatment.	0.56	0.51	–1.40	<b>0.21</b>	–0.90	Keep
49	Prov	My health care provider gave me the correct information about my health situation.	0.67	<b>0.44</b>	–2.20	<b>0.14</b>	–1.90	Drop/revise
51	Prov	I had confidence in my health care provider.	0.59	<b>1.73</b>	<b>3.60</b>	<b>2.08</b>	<b>4.00</b>	Drop/revise
52	Prov	My health care provider was helpful.	0.66	0.60	–1.40	<b>0.23</b>	–1.70	Keep
53	Prov	My health care provider had integrity.	0.64	0.69	–1.00	<b>0.26</b>	–1.50	Keep
54	Prov	My health care provider cared about me.	0.65	<b>0.33</b>	–2.40	<b>0.15</b>	–1.30	Drop/revise
56	Prov	My health care provider was effective.	<b>0.40</b>	<b>2.20</b>	<b>3.60</b>	<b>1.87</b>	1.80	Drop/revise
57	Prov	My health care provider made or kept me healthy.	0.65	0.71	–1.00	<b>0.38</b>	–1.20	Keep
58	P & T	My health care provider used the medical technology accurately.	0.63	<b>1.80</b>	<b>3.50</b>	<b>1.62</b>	<b>2.30</b>	Drop/revise
60	P & T	My health care provider made sure the technology was reliable.	0.75	<b>0.41</b>	–2.30	<b>0.33</b>	–1.20	Drop/revise
62	P & T	My health care provider was knowledgeable about the technology.	0.75	<b>0.41</b>	–2.30	<b>0.19</b>	–1.70	Drop/revise
63	P & T	I trust how my health care provider used the technology.	0.74	<b>0.33</b>	–2.60	<b>0.14</b>	–1.60	Drop/revise
64	P & T	I feel good about the way my health care provider used the technology	0.77	0.57	–1.70	<b>0.37</b>	–1.50	Keep
65	P & T	I have positive feelings about how my health care provider used the technology.	0.77	0.58	–1.80	<b>0.33</b>	–1.80	Keep
68	P & T	My health care provider was successful in their use of the technology.	0.79	<b>0.43</b>	–2.40	<b>0.26</b>	–1.80	Drop/revise
73	P & T	The results my health care provider obtained from the technology were correct results.	0.70	0.89	–0.30	<b>0.42</b>	–0.90	Keep
77	P & T	My health care provider used the technology in a caring manner.	<b>0.38</b>	<b>2.78</b>	<b>4.90</b>	<b>3.37</b>	<b>3.80</b>	Drop/revise
80	P & T	My health care provider had high reliance on the medical technology.	0.60	<b>2.08</b>	<b>4.60</b>	<b>1.99</b>	<b>3.50</b>	Drop/revise

Bold indicates fit statistic that inspired further investigation.

technical qualities as opposed to interpersonal qualities, which is part of an array of research on the trust relationship between physician and patient (Pearson and Raeke, 2000; Fung et al., 2005; Tarn et al., 2005). It is not known if characteristics of technologies used in patient care will affect a patients' trust in medical technologies. The effects of a patient trusting or not trusting medical technologies on patient health outcomes and assessments of quality are also unknown. The validated trust in medical technology instrument, presented in this study, provides a means to explore the role of patient trust in medical technologies in patient–physician relationships.

In a recent comprehensive review (Pearson and Raeke, 2000), researchers describe a need for more research and ways of measuring the relationship between patients and physicians. Insights into the relationship between interpersonal trust and trust in technology in health care work systems can be explored with the instrument describes in this study and will provide an important contribution to health systems research.

## 2. Methods

The trust in medical technology instrument (TMT) was developed using the Rasch instrument development framework and Crocker and Algina (1986) methods as described in (Wolfe and

**Table 7**  
Linacre's rating scale guidelines.

Guideline	Decision
At least 10 observations of each category	Yes
Regular observation distribution	Yes
Average measures advance monotonically with category	Yes
OUTFIT mean-squares less than 2.0	Yes
Step calibrations advance	Yes
Ratings imply measures and measure imply ratings	Yes
Step difficulties advance by at least 1.4 logits	Yes
Step difficulties advance by less than 5.0 logits	Yes

Source: Linacre (2002).

**Table 8**

Rating scale total instrument 1 = disagree, 2 = neutral, 3 = agree.

Total							
Category	Count	Count %	Measure average	Outfit Unweighted MNSQ	Structure measure (t)	Coherence M → C	Coherence C → M
1	441	6	-2.25	1.35	None	64%	24%
2	1165	15	11.16	0.71	-5.59	45%	44%
3	5574	71	30.63	1.32	5.59	88%	93%

Smith, 2007). Items for the instrument were developed by the authors in a series of pilot research studies related to understanding patients' trusting attitudes towards medical technology (Montague et al., 2008, 2009; Montague, 2009). The steps were as follows:

1. Identified that the primary purpose of these measure would be for research purposes, to assess how trust in medical technology may relate to other patient variables in health care work systems.
2. Identified behaviors that represented the construct and defined the domain (Montague et al., 2009).
3. Organized a set of test specifications (Montague et al., 2008).
4. Assembled an initial pool of items (Montague et al., 2009).
5. Pilot testing. Items were reviewed (and revised when necessary) (Montague and Wolfe, unpublished manuscript).
6. Field testing. Held preliminary item tryouts.
7. Determined statistical properties of item scores and, when appropriate, eliminated items that did not meet pre-established criteria.
8. Designed and conducted reliability and validity studies for the final form of the test.
9. Developed guidelines for administration, scoring, and interpretation of the test scores.

The validation study presented in this manuscript describes steps one and six through nine. Steps two through five have already been described in previous publications (Montague et al., 2008, 2009; Montague, 2009). The Crocker and Algina (1986) approach to instrument development was followed and items were developed using methods identified by Jian et al. (2000). The instrument validation process is based on the Messick (1995) validation framework and Rasch validation methods (Wolfe and Smith, 2007).

During the validation phase of instrument development, data were collected from participants with experiences with electronic fetal monitors. To determine the validity of the TMT instrument, analyses were conducted to determine the measure's dimensionality, reliability, fit, and rating scale.

### 2.1. Sample

The TMT instrument was designed in a checklist format, which required the responder to have had a specific experience with the

specified medical technology to complete. Participants were all women who had given birth and used the electronic fetal monitor. Twenty mothers from each age group of 18–35; (i.e. twenty 18 year olds, twenty 19 year olds, twenty 20 year olds, etc.) were randomly invited to complete the instrument through an email data base of women with children to total 360 invited participants. Additionally, an advertisement was placed in an online community for new mothers. One hundred and one participants completed the instrument and some provided additional qualitative responses for each item; data collection was closed after 101 responses were obtained. The average number of children for participants was 1.78. Eighty-four women reported having insurance, 16 women did not have health insurance, and 1 declined to respond to the question. The average age of the participants was 25.28, and the average years of education was 14.53. Participants identified as white or Caucasian ( $n = 82$ ), African, black, African American or Caribbean ( $n = 8$ ), Asian ( $n = 3$ ), Hispanic ( $n = 1$ ) or declined to select a racial or ethnic group ( $n = 7$ ). To have a representative range of childbirth experiences and levels of trust with technology, participants were recruited who used the electronic fetal monitors in a variety of contexts including home births, in hospitals, intermittently, and continuously. Participants also used a variety of primary care providers for their births including physicians, registered nurses, nurse midwives, doulas, and rotating physicians.

### 2.2. Measurement model

Item Response Theory (IRT), sometimes referenced as latent trait theory, is a measurement model that is an alternative to true score test theory. IRT assumes a link between a participant's response on a test item and the construct being measured (trust in medical technology). IRT makes stronger assumptions than classical test theory and in many cases provides correspondingly stronger findings (Kline, 2005). IRT provides several improvements in scaling items and people (Hambleton et al., 1991) and is better at predicting item bias amongst different groups, such as gender and ethnicity (Hogan, 2003). One set back of using this method is that IRT requires complicated estimations when not using the basic Rasch model (Hambleton et al., 1991). IRT was used with the Rasch Rating Scale model in this study.

The Rasch Rating Scale model was used in this study, because the TMT collects Likert-type responses and the models require smaller sample sizes for validation (Wolfe and Smith, 2007). As illustrated in the equation below, the model specifies that the probability that

**Table 9**

Rating scale for scale 1 technology 1 = disagree, 2 = neutral, 3 = agree.

Scale 1							
Category	Count	Count%	Measure average	Outfit Unweighted MNSQ	Structure measure (t)	Coherence M → C	Coherence C → M
1	289	10	-8.65	1.15	None	69%	39%
2	653	24	7.48	0.8	-8.81	53%	60%
3	1546	56	26.34	1.46	8.81	84%	86%

**Table 10**  
Rating scale for scale 2 technology 1 = disagree, 2 = neutral, 3 = agree.

Scale 2								
Category	Count	Count%	Measure average	Outfit Unweighted MNSQ	Structure measure (t)	Coherence M → C	Coherence C → M	
1	96	5	−6.6	1.15	None	51%	19%	
2	235	13	12.94	0.56	−6.22	48%	51%	
3	1514	81	35.26	1.83	6.22	92%	95%	

a respondent will respond to a particular item is a logistic function of the respondent's level of the attitude ( $\theta$ ) and the item's representation of the construct ( $\delta$ ). The probability ( $\pi_{nix}$ ) that a specific respondent  $n$  will rate a particular item  $i$  with a specific rating scale category  $x$ . The model includes a threshold difficulty parameter ( $\tau_k$ ) that depicts the difficulty of moving from one scoring category to another on a polytomous (multiple response option) item. It is assumed that the distance between each category threshold is constant across items within the same rating scale. The WINSTEPS (Linacre, 2002) software package was used to estimate the parameters in the model, which were reported on a single linear continuum in logistic odds ratio units (logits) (Smith, 2000).

$$\pi_{nix} = \frac{\exp \sum_{j=0}^x D(\theta_n - \delta_i - \tau_k)}{\sum_{k=0}^m \exp \sum_{j=0}^k D(\theta_n - \delta_i - \tau_j)}$$

### 2.3. Dimensionality

A principal component analysis was performed to evaluate the dimensionality of the instrument. The residual components' eigenvalues and the percentage variance accounted for by each component were calculated. In congruence with Kaiser's Criterion, only components with an eigenvalue greater than 1.00 were retained and components with four or more loadings with absolute values greater than 0.60 were selected as reliable components for interpretation (Kaiser, 1970; Kline, 2005). Items with absolute value loadings greater than .30 were considered part of the construct. Each dimension was then interpreted based on apparent similarities in the content of items loaded on that dimension. Principal components analysis and scree were performed to make sure the model's assumption of a unidimensional latent trait were not violated.

### 2.4. Reliability

Person and item reliability were conducted in WINSTEPS. In WINSTEPS person reliability is the same as test reliability. Low values in person reliability indicate a tight range of person measures, or an undersized number of items; low item reliability scores indicate a narrow range of item measures, or a small sample size (Linacre, 2002). High reliability (of persons or items) means that there is a high probability that persons (or items) estimated with high measures actually do have higher measures than persons (or items) estimated with low measures (Linacre, 2002).

### 2.5. Fit

Item fit indices were calculated to confirm whether the items identify the variable targeted for measurement (trust in medical technology). Person fit indices verified whether participants

responded to the items in the way they are supposed to, as depicted by the model. Item and person fit indices were calculated using WINSTEPS computer software (Linacre, 2002).

First point-measure correlations (PTMEA CORR.),  $r_{pm}$  or RPM, were examined between the observations on an item and the persons estimated level of trust (Linacre, 2002). Point-measure correlations over 0.5 were considered good, while 0.3 or 0.4 required scrutiny of the item. In the weighted infit mean-square statistic (INFIT) MNSQ column values less than two, but close to one were examined and items above 1.5 were considered too high. Infit MNSQ are standard fit statistics based on chi-square statistics, which are more sensitive to patterns in observations. Outfit z-standardized (OUTFIT) ZSTD fit statistics are based on chi-square statistics and are more sensitive to unexpected observations by persons on items (Linacre, 2002). Unweighted (OUTFIT) ZSTD is considered standard, if the absolute value is greater than two the item was flagged for further examination (Smith, 2000). Items with unweighted MNSQ items under one were ignored, because that score indicated over fit, which meant the model was doing a better job at predicting than possible. Values around 1.4 were used, while values greater than 1.4 were scrutinized. When items were scrutinized, items were reassessed qualitatively for discrepancies such as the wording, format, or item position that may cause the item to function differently than other items.

### 2.6. Rating scale analysis

A rating scale analysis was conducted to examine the use of the rating scale categories for polytomous items. The analysis provided additional information about whether or not the structure the TMT required respondents to utilize when providing responses was employed in the manner in which it was intended (Wolfe and Smith, 2007). Generalized p-values indicated the degree to which the answers for the item are difficult, relative to the total available points (Wolfe and Smith, 2007). Point-polyserial, polyserial, and point-measure correlations indicated the degree to which the item scores were consistent with the total test scores while fit indices indicated the degree to which the scored responses of individual respondents were consistent with the expectations of the Rasch model (Wolfe and Smith, 2007). For methodical purposes the goal was to verify that the rating scale observation conform closely to a specified model (Linacre, 2002).

## 3. Results

### 3.1. Dimensionality

The principal components analysis provided evidence for multidimensionality, which lead to the creation of three subscales. Kaiser's rule indicated 13 possible subscales, but only three possessed enough items and logic to be included. Eigenvalues were calculated, variance was accounted, and Kaiser's criterion was used to determine the numbers of components retained in the instrument (see Table 2). The first factor's eigenvalue is 31.17 and

accounts for 39% of the variance. The eigenvalues of two or more components are well above Kaiser's criterion; suggesting retention of three or more components and thus multidimensionality. The three largest components represent the three domain areas represented in the instrument design model; technology, provider and how the provider uses the technology.

To ensure satisfaction of the unidimensional assumption of the Rasch rating scale model, item analyses were performed on subscales separately. An investigation of the item loadings of the first residual show that all of the items on the factor are positively related to trust in technology (see Table 3).

Factor two has positive loadings on items related to the health care provider (see Table 4).

Factor three positively loading items are related to how the provider uses the technology (see Table 5).

### 3.2. Fit

Point-measure correlations generally remained in the expected range. 39 items were flagged for further evaluation because of mean-square weighted (MS),  $z$  weighted ( $Z_w$ ), mean-square unweighted ( $MS_u$ ), or  $z$  unweighted ( $Z_u$ ) scores above two (see Table 6). Criterion for evaluation was based on Rasch validation guidelines. Point-measure correlations over 0.5 are considered good, while 0.3 or 0.4 require scrutiny of the item (Wolfe and Smith, 2007). Mean-square weighted and unweighted statistic ( $MS_w$  and  $MS_u$ ) outside of the 0.5–1.5 range of productive of measurement were also flagged for further evaluation (Linacre, 2004). Weighted and unweighted mean-square fit statistics between  $-2$  and  $2$  were considered to be functioning properly (Smith, 2000). Items with three or more flags should be excluded from use as indicated in Table 6.

### 3.3. Rating scale analysis

Items in all four analyses, the total instrument, scale 1, scale 2, and scale 3, appear to be functioning well for the three scale points agree, neutral and disagree based on plots of category probabilities and Linacre's rating scale guidelines (see Table 7).

The preliminary guideline is that all items orientate with the latent variable (Linacre, 2004). All items employed the same rating scale and therefore cooperated to a shared latent variable. Rating scale categories for negatively-oriented items often function differently in rating scale categories (Yamaguchi, 1997). A separate study generated factors of the construct trust in medical technology and found trust and distrust to be theoretical opposites, which provides more evidence that all items are orientated with the latent variable (Montague et al., 2009).

Guideline #1 is that at least 10 observations must occur in each category (Linacre, 2004). The lowest number of observations for the total instrument and separate sub scales is 55, which is well above the 10-observation guideline, therefore meeting guideline #1 for the total instrument and subscales (see Tables 8–11).

Guideline #2 is that there must be regular observation distribution (Linacre, 2004). It was expected that the population would

be mostly trusting (Anderson and Dedrick, 1990; Thom et al., 1999). Therefore the distribution count should have higher number of counts for 3 = agree, which indicated higher trust and descend to category 1 = disagree. This distribution pattern is consistent across the total instrument and all three subscales (see Tables 8–11).

Guideline #3 is that average measures must advance monotonically with category (Linacre, 2004). Smith (2000) says, "observations in higher categories must be produced by higher measures; which means that the average measures by category, for each empirical set observations, must advance monotonically up the rating scale," (p. 22). This guideline was met by measure averages across the four analyses (see Tables 8–11). This means that empirically category 3 (trust) represents a higher level of the construct trust than categories two (no trust) and three (distrust).

Guideline #4 is that OUTFIT mean-squares must be less than 2.0 (Linacre, 2004). All unweighted mean-squares are less than 2.0 in both the total instrument and subscales, meeting the requirement for guideline 4.

Guideline #5 is that step calibrations must advance, which is an indication of the rating scales ability to increase proportionately with the level of the construct (Linacre, 2004). The "scale structure measures", also called "step calibrations", progress in all occurrences, meeting the requirements for guideline 5 (see Tables 8–11).

Guideline #6 is that ratings imply measures and measure imply ratings (Linacre, 2004). This guideline says that the rating should imply the measure and vice versa, this is determined with the coherence statistic (see Tables 8–11). The computation coherence is outlined in Table 6 as  $(X)M \rightarrow C$  (Measure implies Category %) which indicates the percentage of the rating that is expected to be observed in a category and are actually observed in the category (Linacre, 2004). This guideline was met and is indicated by moderate coherence statistics.

Guideline #7 and 8 states that step difficulties must advance by at least 1.4 logits and less than 5.0 logits (Linacre, 2004). Scale structure measures ( $t$ ) represent step difficulties and advance a distance of 1.7 logits for all categories in the rating scale which is between 1.4 and 5.0 logits (see Tables 8–11).

### 3.4. Reliability

The TMT produced high reliability across the various subscales for person (test) and item reliability ranging from 0.71 to 0.92 (see Table 12).

## 4. Discussion

Our results indicate that the trust in medical technology instrument is sufficiently valid and ready for use. Future studies will involve validation in a variety of medical domains including consumer health products. The validation studies found that trust in medical technology is in fact a multidimensional construct involving subscales of technology characteristics, provider characteristics and characteristics of the how the provider uses technology. When disseminating this instrument, subscales can be used

**Table 11**  
Rating scale for scale 3 technology 1 = disagree, 2 = neutral, 3 = agree.

Scale 3							
Category	Count	Count%	Measure average	Outfit Unweighted MNSQ	Structure measure ( $t$ )	Coherence M $\rightarrow$ C	Coherence C $\rightarrow$ M
1	55	5	-12.8	1.5	None	75%	43%
2	277	23	13.2	0.69	-16.66	66%	55%
3	826	69	38.66	1.15	16.66	86%	93%

**Table 12**

KR-20, real and modeled person reliability, and real and modeled item reliability for scales 1–3 and total instrument.

Scale	Items	Winsteps (KR-20)	Person reliability (real)	Person reliability (modeled)	Item reliability (real)	Item reliability (modeled)
total	80	1*	0.86	0.87	0.92	0.92
1	31	1*	0.81	0.84	0.82	0.86
2	26	0.96*	0.53	0.56	0.86	0.87
3	23	1	0.58	0.58	0.71	0.75

individually or collectively, as each subscale was validated individually and as a complete instrument.

The purpose of the fit indices was to flag potentially problematic response patterns. Once a flag has been raised, it is the responsibility of the data analyst or instrument developer to seek plausible explanations for the flag. It is unwise to simply delete an item from an instrument based on a fit flag alone. These indices provide relative measures and are subject to Type I statistical errors. By scrutinizing the misfitting items or persons or the patterns of responses associated with them, valuable knowledge may be gained. Misfit might occur for persons because of guessing, familiarity with the test, carelessness, or because the user has specialized knowledge or deficiencies. Misfitting may occur when there is multidimensionality, the quality of the items is poor, the item alludes to a correct answer, or because of miskeying. Therefore, items flagged as misfitting should be evaluated and redesigned for future validation studies (Wolfe and Smith, 2007). The flagged items should be used in future validation studies and reassessed for fit. In general use, items with fewer than two flags should be reworded and reassessed for fit as indicated in Table 6.

All of Linacre's (2004) guidelines were met, which provided evidence of measure stability, measure accuracy (fit), description of the sample, and inference for the next sample (see Table 7). Additionally, the instrument produced high reliability across the various subscales for person (test) and item reliability, using both Cronbach's alpha and the more conservative KR-20. Reliability scores indicate high internal consistency of the total instrument and subscales. A rating scale analysis was conducted to assess the appropriate rating scale for the instrument using Linacre's framework for rating scale evaluation (Linacre, 2004). Each of the guidelines was met for rating scale analysis and therefore the rating scale is appropriate for the instrument (Table 13).

**Table 13**

Summary of guideline pertinence.

Guideline	Measure stability	Measure accuracy	Description of sample	Inference for next sample
Pre. Scale oriented with latent variable	Essential	Essential	Essential	Essential
At least 10 observations of each category	Essential	Helpful		Helpful
Regular observation distribution	Helpful			Helpful
Average measures advance monotonically with category	Helpful	Essential	Essential	Essential
OUTFIT mean-squares less than 2.0	Helpful	Essential	Helpful	Helpful
Step calibrations advance			Helpful	
Ratings imply measures and measure imply ratings				Helpful
Step difficulties advance by at least 1.4 logits		Helpful		Helpful
Step difficulties advance by less than 5.0 logits	Helpful			

#### 4.1. Validity evidence

Messick (1995) proposed a six-component framework for validity assessment involving content, substantive, structural, generalizability, external and consequential aspects of validity (Messick, 1995). Validity evidence was provided through test (Table 14) and data based evidence (Table 15).

The content aspect refers to evidence of content relevance, representativeness, and technical quality of the instrument. The documentation of the purpose and use of the instrument, the domain, the development of the test blueprint, documentation of the test development process and expert review of the instrument measure content aspects of validity. These steps were conducted in studies prior to this validation study (Montague et al., 2008, 2009; Montague, 2009)

The substantive aspect refers to the theoretical rationale for observed consistencies in responses, including process models of task performance and empirical evidence that the theoretical

**Table 14**

Data based validity evidence

Aspect of validity	Validity evidence
Structural	<ul style="list-style-type: none"> <li>✓ Demonstrating internal relationships among items (structural fidelity) or subsets of items consistent with the underlying theory (factor analysis, item difficulty hierarchies that are consistent with the construct map)</li> <li>✓ Demonstrating that conditional difficulties of items are equal or consistent with known influences on test performance (differential item functioning analysis)</li> <li>✓ Utilizing a measurement model that combines (weights) information across observations (scales) and takes into account (controls for) undesirable influences on the scores in a manner that is consistent with intended score interpretation (e.g., criterion-referenced versus norm-referenced).</li> </ul>
Generalizability	<ul style="list-style-type: none"> <li>✓ Demonstrating reliability of measures from the instrument across a variety of contexts                             <ul style="list-style-type: none"> <li>✓ Internal consistency</li> <li>× Test-retest</li> <li>× Interrater</li> <li>× Alternate form</li> <li>× Application of meta-analytic procedures to validity coefficients across a variety of measurement contexts and samples (validity generalization)</li> </ul> </li> </ul>
External	<ul style="list-style-type: none"> <li>× Demonstrating that predicted group differences are realized empirically</li> <li>× Experimentation: Any theory-based comparison that yields outcomes consistent with theory.</li> <li>× Comparison of Groups: Differences between groups that theory predicts will be different on the underlying construct</li> <li>× Changes Over Time: Individuals are typically expected to change over time as a result of maturation. Observed differences such as these can also be used as evidence for construct validity.</li> </ul>
Consequential	<ul style="list-style-type: none"> <li>✓ Detecting positive or negative impact on individuals (bias) (e.g., identifying substantively explainable DIF versus sources of item or test bias, determining whether the labels generated in test use result in stigmatization of some groups, expert judgment about the suitability of test content for making decisions about individuals and groups)</li> </ul>

**Table 15**  
Test based validity evidence.

Aspect of validity	Validity evidence
Content	<ul style="list-style-type: none"> <li>✓ Documentation of purpose and uses of the instrument</li> <li>✓ Documentation of use of domain analysis</li> <li>✓ Development of test blueprint &amp; item templates</li> <li>✓ Documentation of test development process</li> <li>✓ Expert Review (Readability, Clarity, Clarity of instructions, Fairness, Sensitivity)</li> </ul>
Substantive	<ul style="list-style-type: none"> <li>✓ Development of operationalized definition &amp; theoretical framework of the construct (including internal, external, &amp; processing models)</li> <li>✓ Documentation of use of expertise</li> <li>× Verifying of use of the proposed processes by respondents (think aloud)</li> </ul>
Structural	<ul style="list-style-type: none"> <li>✓ Development of internal model</li> <li>✓ Rationale for adoption of reference framework (criterion versus normative)</li> <li>✓ Rationale for developing response format (rating scale and/or distracters)</li> <li>✓ Rationale for adoption of measurement model</li> </ul>
Generalizability	<ul style="list-style-type: none"> <li>✓ Specification of target population</li> <li>✓ Selection of norming population</li> </ul>
External	<ul style="list-style-type: none"> <li>✓ Demonstrating that the construct theory is sufficiently developed to support investigations of external aspect evidence</li> </ul>
Consequential	<ul style="list-style-type: none"> <li>× Detecting positive or negative impact on systems (systemic validity)</li> </ul>

processes are actually engaged by respondents in the assessment tasks. Structural aspects of validity refer to the fidelity of the scoring structure to the structure of the construct domain. Activities towards assessing structural validity include developing an internal model of the construct; the model was developed from a review of the literature and evaluated with the dimensionality analysis. Evidence of structural validity was also based on a rationale for the reference and response framework and measurement model, which were evaluated by the rating scale analysis.

The generalizability aspect refers to the degree to which test score properties and interpretations generalize to and across population groups, settings, and tasks including validity generalization of test criterion relationships. This was established by specifying the target population for the instrument (patients in health care work systems) and a norming population (a diverse set of obstetric patients).

External aspects of validity include convergent and discriminate evidence from multitrait-multimethod comparisons as well as evidence of criterion relevance and applied utility. Consequential aspects of validity refers to the value implications of score interpretation as a basis for action as well as the actual and potential consequences of test use, especially regarding bias, fairness, and distributive justice. Analyzing the suitability of the test content for individuals and groups that could potentially be stigmatized by scores assessed consequential aspects of validity.

The table below illustrates efforts that were made to contribute to the various aspects of validity with test based and data based evidence; checks represent evidence of validity that were found in the current study, X's represent aspect that should be evaluated in future validation efforts (Wolfe and Smith, 2007).

#### 4.2. Limitations of the study

Instrument validation often occurs over a period of time over several validation studies. This instrument was tested with only one patient population, as is appropriate for the assessment of

validity (Kline, 2005). Future studies should validate the instrument with a larger sample of participants and in different health settings. The response rate for the instrument was under 30%, which may be a sign of response bias; future research should explore the characteristics of participants who choose not to participate in trust related studies. Some aspects of validity were not assessed in this study and should be addressed in future studies to strengthen the validity assessment of the instrument. Substantive validity could be further explored by verifying the use of the proposed processes by respondents using think aloud. Consequential validity could be assessed, by detecting the positive or negative impact on the system; this would answer the question of how understanding and/or measuring trust in medical technology will impact how systems function over time. Generalizability could be improved by testing the instrument in other contexts, using test-rest, inter-rater options, alternate forms, and applying a meta analytic procedure to validate coefficient across a variety of measurement contexts and samples, such as other health domains, and a more diverse sample of users (i.e. other cultures, backgrounds, levels of illness). External validity could also be further explored with future studies that explore hypothesized differences across groups. For example studies have shown that minorities have a higher degree of distrust in health systems and physicians. Future research can explore racial and ethnic disparities in the context of medical technology (Rose et al., 2004). Individual differences in trust in medical technology may exist in relation to positive or negative medical outcomes and the amount of time between the medical event and measuring the construct. Future studies should also explore the relationship of trust in medical technology through experimentation, any theory based comparison that yields outcomes consistent with theory. Another aspect of validity that was not explored in this study was user changes over time. Individuals are typically expected to change over time as a result of maturation; these observed differences can be used as evidence of construct validity. The present validation will help understand patient trust in medical technologies, which is a gap in the present literature. The authors hope that this work will inspire interest in the topic and that the validated instrument will support future research on the subject.

#### 5. Conclusion

An instrument was developed to assess patients' trusting attitudes towards medical technologies. Validity was assessed using Messick's (1995) framework; support towards instrument validity was presented in the results in the form of both test and data based evidence for all but consequential and external aspects of validity. The instrument measures patient trust in medical technology using three subscales, in accordance with findings of multidimensionality of the construct. High reliability scores across the various subscales for person (test) and item reliability are evidence of instrument internal consistency.

#### Acknowledgements

This work was partially supported by the Francis Research Fellowship and by grant 1UL1RR025011 from the Clinical and Translational Science Award (CTSA) program of the National Center for Research Resources, National Institutes of Health. The authors are grateful for comments about the development of the instrument provided by Dr. Brian Kleiner, Dr. Woodrow Winchester III, Dr. Tonya Smith-Jackson and Dr. Bernice Hausman. The authors also appreciate comments and suggestions from the editor Dr. Ann Bisantz and two anonymous reviewers.

## References

- Anderson, L.A., Dedrick, R.F., 1990. Development of the trust in physician scale: a measure to assess interpersonal trust in patient–physician relationships. *Psychological Reports* 67 (3.2), 1091–1100.
- Awad, N.F., Ragowsky, A., 2008. Establishing trust in electronic commerce through online word of mouth: an examination across genders. *Journal of Management Information Systems* 24 (4), 101–121.
- Basso, A., Goldberg, D., Greenspan, S., Welmer, D., 2001. First impressions: Emotional and cognitive factors underlying judgments of trust e-commerce. *Proceedings of the Third ACM Conference on Electronic Commerce*, pp. 137–143.
- Bisantz, A.M., Seong, Y., 2001. Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics* 28 (2), 85–97.
- Boehm, F.H., 2003. Building trust. *Family Practice News* 33 (15), 12.
- Buckle, P., Clarkson, P.J., Coleman, R., Ward, J., Anderson, J., 2006. Patient safety, systems design and ergonomics. *Applied Ergonomics* 37 (4), 491–499.
- Calnan, M., Montaner, D., Horner, R., 2005. How acceptable are innovative health-care technologies? A survey of public beliefs and attitudes in England and Wales. *Social Science & Medicine* 60 (9), 1937–1948.
- Castelfranchi, C., Falcone, R., 2002. Social trust: a cognitive approach. *Trust and Deception in Virtual Societies*: pp. 55–90.
- Corbitt, B.J., Thanasankit, T., Yi, H., 2003. Trust and e-commerce: a study of consumer perceptions. *Electronic Commerce Research and Applications* 2 (3), 203–215.
- Crocker, L., Algina, J., 1986. *Introduction to Classical and Modern Test Theory*. Holt, Rinehart and Winston.
- DeFuria, G.L., 1996. *Interpersonal Trust Surveys*. Jossey-Bass, San Francisco.
- Dzindolet, M.T., Peterson, S.A., Pomranky, R.A., Pierceb, L.G., Beck, H.P., 2003. The role of trust in automation reliance. *International Journal of Human-Computer Studies* 58, 697–718.
- Franks, P., Fiscella, K., Shields, C.G., Meldrum, S.C., Duberstein, P., Jerant, A.F., Tancredi, D.J., Epstein, R.M., 2005. Are patients' ratings of their physicians related to health outcomes. *Annals of Family Medicine* 3 (3), 229–234.
- Fung, C.H., Elliott, M.N., Hays, R.D., Kahn, K.L., Kanouse, D.E., McGlynn, E.A., et al., 2005. Patients' preferences for technical versus interpersonal quality when selecting a primary care physician. *Health Serv Res* 40 (4), 957–977.
- Grabner-Krauter, S., Kaluscha, E.A., 2003. Empirical research in on-line trust: a review and critical assessment. *International Journal of Human-Computer Studies* 58, 783–812.
- Groeneveld, P.W., Sonnad, S.S., Lee, A.K., Asch, D.A., Shea, J.E., 2006. Racial differences in attitudes toward innovative medical technology. *Journal of General Internal Medicine* 21 (6), 559.
- Hambleton, R.K., Swaminathan, H., Rogers, H.J., 1991. *Fundamentals of Item Response Theory*. Sage, Newbury Park.
- Hoc, J., 2000. From human-machine interaction to human-machine cooperation. *Ergonomics* 43 (7), 833–844.
- Hogan, T.P., 2003. *Psychological Testing: A Practical Introduction*. John Wiley and Sons, New York.
- Jian, J.Y., Bisantz, A.M., Drury, C.G., 2000. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics* 4 (1), 53–71.
- Kjerulf, K.H., Pillar, B., Mills, M.E., Lanigan, J., 1992. Technology anxiety as a potential mediating factor in response to medical technology. *Journal of Medical Systems* 16 (1), 7–13.
- Kaiser, H.F., 1970. A second generation little jiffy. *Psychometrika* 35, 401–415.
- Kao, A.C., Green, D.C., Davis, N.A., Koplan, J.P., Cleary, P.D., 1998. Patients' trust in their physician: effects of choice, continuity, and payment method. *Journal of General Internal Medicine* 13 (10), 681–686.
- Kline, T.J.B., 2005. *Psychological testing: a practical approach to design and evaluation*. Sage, Thousand Oaks.
- Kong, W., Hung, Y., 2006. Modeling initial and repeat online trust in b2c e-commerce. In: *Hawaii International Conference on System Sciences*. IEEE, p. 120.
- Larzelere, R.E., Huston, T.L., 1980. The dyadic trust scale: toward understanding interpersonal trust in close relationships. *Journal of Marriage and the Family* 42 (3), 595–604.
- Lee, J.D., Moray, N., 1994. Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies* 40, 153–184.
- Lee, J.D., See, K.A., 2004. Trust in automation: designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46, 50–80.
- Lewandowsky, S., Mundy, M., Tan, G.P.A., 2000. The dynamics of trust: comparing humans to automation. *Journal of Experimental Psychology: Applied* 6 (2), 104–123.
- Linacre, J.M., 2002. *Winsteps*. 3.63 ed. Chicago.
- Linacre, J.M., 2004. *Optimizing Rating Scale Category Effectiveness*, Introduction to Rasch Measurement Maple Grove, Minnesota.
- Martin, J.L., Norris, B.J., Murphy, E., Crowe, J.A., 2008. Medical device development: the challenge for ergonomics. *Applied Ergonomics* 39 (3), 271–283.
- McKnight, D.H., Choudhury, V., 2006. Distrust and trust in b2c e-commerce: do they differ?. In: *Proceedings of the Eighth International Conference on Electronic Commerce: The New E-commerce: Innovations for Conquering Current Barriers, Obstacles and Limitations to Conducting Successful Business on the Internet*. ACM New York, NY, USA, pp. 482–491.
- McKnight, D.H., Choudhury, V., Kacmar, C., 2003. Developing and validating trust measures for e-commerce: an integrative typology. *Information Systems Research* 13 (3), 334–359.
- Messick, S., 1995. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist* 50 (9), 741–749.
- Moffa, A.J., Stokes, A.F., 1997. Trust in a medical expert system: can we generalize between domains. In: Mouloua, M., Koonce, J.M. (Eds.), *Human-Automation Interaction: Research and Practice*. Lawrence Erlbaum Associates, Mahwah, New Jersey.
- Montague, E., 2009. Patient user experience of medical technology. In: *Ergonomics and Health Aspects of Work with Computers*. Lecture Notes in Computer Science 5624. Springer, pp. 70–77.
- Montague, E., Kleiner, B.M., Winchester, W.W., 2009. Empirical evaluation of the construct trust in medical technology. *International Journal of Industrial Ergonomics* 39 (4), 628–634.
- Montague, E., Wolfe, E.W., Unpublished manuscript. Preliminary development of a trust in medical technology instrument.
- Montague, E., Wolfe, E.W., Kleiner, B.M., Winchester, W.W., 2008. Using Rasch Analysis to Construct a Trust in Medical Technology Instrument. Institute of Objective Measurement Workshop, New York, NY.
- Moray, N., 2000. Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied* 6 (1), 44–58.
- Muir, B., 1994. Trust in automation: part 1. Theoretical issues in the study and human intervention in automated systems. *Ergonomics* 37, 1905–1923.
- Muir, B., Moray, N., 1996. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* 39 (3), 429–460.
- Parasuraman, R., Riley, V., 1997. Humans and automation: use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39, 230–253.
- Pearson, S.D., Raeke, L.H., 2000. Patients' trust in physicians: many theories, few measure, and little data. *Journal of General Internal Medicine* 15, 509–513.
- Rempel, J.K., Holmes, J.G., Zanna, M.P., 1985. Trust in close relationships. *Journal of Personality and Social Psychology* 49 (1), 95–112.
- Rose, A., Peters, N., Shea, J.A., Armstrong, K., 2004. Development and testing of the health care system distrust scale. *Journal of General Internal Medicine* 19, 57–63.
- Sheridan, T.B., 2002. *Humans and Automation*. John Wiley and Sons, Santa Monica.
- Smith, R.M., 2000. Fit analysis in latent trait measurement models. *Journal of Applied Measurement* 1 (2), 199–218.
- Tarn, D.M., Meredith, L.S., Kagawa-Singer, M., Matsumura, S., Bitto, S., Oye, R.K., Liu, H., Kahn, K.L., Fukubara, S., Wenger, N.S., 2005. Trust in one's physician: the role of ethnic match, autonomy, acculturation, and religiosity among Japanese Americans. *Annals of Family Medicine* 3 (4), 339–347.
- Thom, D.H., Ribisl, K.M., Stewart, A.L., Luke, D.A., 1999. Further validation and reliability testing of the trust in physician scale. *Med Care* 37 (5), 510–517.
- Thom, D.H., Kravitz, R.L., Bell, R.A., Krupat, E., Azari, R., 2002. Patient trust in physician: relationship to patient requests. *Family Practice* 19 (5).
- Timmons, S., Harrison-Paul, R., Crosbie, B., 2008. How do lay people come to trust the Automatic external defibrillator? *Health Risk & Society* 10 (3), 207–220.
- Tschannen-Moran, M., Hoy, W.K., 2000. A multidisciplinary analysis of the nature, meaning, and measurement of trust. *Review of Educational Research* 70 (4), 547–593.
- Wolfe, E.W., Smith, E.V., 2007. Instrument development tools and activities for measure validation using Rasch models: part II validation activities. In: Smith, E.V., Smith, R.M. (Eds.), *Rasch Measurement: Advanced and Specialized Applications*. JAM Press, Maple Grove, MN.
- Yamaguchi, J., 1997. Positive vs negative wording. *Rasch Measurement Transactions* 11, 567.
- Zheng, B., Hall, M.A., Dugan, E., Kidd, K.E., Levine, D., 2002. Development of a scale to measure patients' trust in health insurers. *Health Services Research* 37 (1), 187–202.