International Journal of Engineering Education Vol. 37, No. 3, pp. 733–743, 2021
Printed in Great Britain
0949-149X/91 $3.00+0.00

Measuring Empathy for Users in Engineering Design*

JUSTIN L. HESS
School of Engineering Education, Purdue University, West Lafayette, IN, USA. E-mail: jhess@purdue.edu

NICHOLAS D. FILA
Electrical and Computer Engineering, Iowa State Univeristy, Ames, Iowa, USA. E-mail: nfila@iastate.edu

EUNHYE KIM
School of Engineering Education, Purdue University, West Lafayette, IN, USA. E-mail: kim1906@purdue.edu

SENAY PURZER
School of Engineering Education, Purdue University, West Lafayette, IN, USA. E-mail: purzer@purdue.edu

This study describes the design and testing of an instrument that measures engineering student tendencies to empathize with users in the context of engineering design. The instrument design seeks to measure three empathy types: Affective Empathy, Imagine-Self Perspective-Taking, and Imagine-Other Perspective Taking. Moreover, the survey includes three sections wherein students respond to how they utilized empathy distinctly in three phases of engineering design: (1) Needfinding, (2) Concept Generation, and (3) Evaluation. We performed confirmatory factor analyses on two distinct construct configurations. First, we created measurement models that sought to measure empathy types without accounting for design phase. Second, we created measurement models that accounted for how these empathy types might manifest distinctly by design phase. We were able to achieve robust measurement models in both configurations, but the set that accounted for design phase enabled the retention of all survey items, thus suggesting the importance of accounting for how these empathy types manifest distinctly across design. However, at this stage of survey development, measuring empathy types by design phases still poses internal consistency concerns due to a limited number of items. Future work involves expanding the set of items for each empathy type within each design phase, distinguishing between affective empathy types, and building on qualitative data to ensure the instrument covers all aspects of students' design experiences wherein empathy manifests.

Keywords: empathy; design; instrument design; validation

1. Introduction

Empathy enables engineering designers to connect with and accurately identify the needs of users and broader stakeholders. Thus, empathy plays an important role in engineering design [1] and is one predictor of innovative behaviors [2]. There is a growing interest in understanding the role of empathy in engineering work [3, 4]. However, previous studies demonstrate limitations in empathic tendencies of engineering students, including student perceptions that empathy does not interface with engineering work [5, 6]. Hence, there is also a growing interest among educators who aspire to promote empathic competencies as part of engineering education [7].

Many well-established instruments exist to explore general empathic tendencies, beliefs, and behaviors [8], but engineering practice provides a unique context and lens of empathic development and application [3, 9]. Moreover, empathy is a complex construct with many related (but distinctly measurable) phenomena [10]. Thus, educators and researchers need an instrument that accounts for the distinct contexts experienced by engineering students and the ways empathy manifests in those contexts.

2. Study Purpose

The objective of this study was to design and ascertain the structural validity of an engineering-specific measure of empathy that was applicable for use in design contexts. The following Research Questions (RQs) guided this investigation:

RQ1: What potential set of survey items align with a four-part empathy model that accounts for how these empathy types manifest across three engineering design phases?
RQ2: To what extent are empathy constructs structurally stable when tested via confirmatory factor analysis?
RQ3: To what extent are factor structures internally consistent?

This study begins with an overview of the instrument design process. First, we present a construct definition that builds on a four-part empathy model

* Accepted 22 December 2020.

and map a series of research-based individual items to the constructs in the four-part empathy model. Second, we describe the research methods, including the data collection process and confirmatory factor analysis procedures. Third, we provide results aligned with two distinct measurement models. Fourth, we provide an overview of the internal consistency of constructs identified through CFA. Finally, we discuss these findings, including potential considerations for immediate use of this instrument as well as future work needed to improve upon its design.

3. Conceptual Model & Instrument Design

To guide the instrument design, we employed an empathy model comprised of four empathy types (see Fig. 1). Like the Interpersonal Reactivity Index [11], these four empathy types vary along two continua: self/other orientation and cognitive/affective emphasis. These four empathy types represent four of eight common uses of the term empathy described by Batson [10] and a sub-set of the facets of the empathy in engineering model described by Walther et al. [3]. Thus, this four-part empathy model does not provide an exhaustive representation of empathy. Nonetheless, the parsimony of the four-part model makes it a useful heuristic for measuring this complex phenomenon.

Affective empathy types represent feelings or emotions which may manifest automatically or within the subconscious. We describe *other-oriented affective empathy* as **empathic concern**, or the tendency to have feelings of concern *for* another or others, especially in recognition of their hardships or suffering. We describe *self-oriented affective empathy* as **empathic distress**, or the tendency to internalize feelings of concern or joy resulting from considerations or interactions with others. Hoffman [12] conceived of empathic distress as a critical driver of prosocial behavior; however, we recognize that other 'self-oriented affective' empathy types exist, such as *emotion congruence* or *motor mimicry* [10].

Cognitive empathy types represent more 'advanced' forms of empathy [12] and tend to require some level of rational thought. We defined *self-oriented cognitive empathy* as **imagine-self perspective-taking**, or the tendency to imagine one's self in another's shoes. This empathy type is similar but slightly distinct from another self-oriented cognitive empathy type which Batson [10] described as *projection* or *einfuhlung* – the primary distinction is that these phenomena also emphasize the importance of imagining how one might feel in another's shoes while encountering a specific situation. Finally, we describe *other-oriented cognitive empathy* as **imagine-other perspective-taking**, or the tendency to imagine another in their own shoes. In brief, the imagine-other and imagine-self perspective-taking distinctions represent the distinction between me imagining myself in your shoes versus me imagining you in your own shoes.

While these empathy types manifest discretely, they can inform each other, often in subconscious and automatic ways. We describe *pluralism* as the shifting between self/other orientations [12] and *mode switching* [3] as shifting between affect/cognitive empathy types. Moreover, like Oxley [13], we argue that "true empathy" involves some level of manifestation of each of these four empathy types. For example, one can consider the perspective of another but have no level of affective relationality; such an encounter is common among sociopaths, who may be very good at considering other's perspectives but seek egocentric gain therefrom [14].

The four empathy types guided the design of the

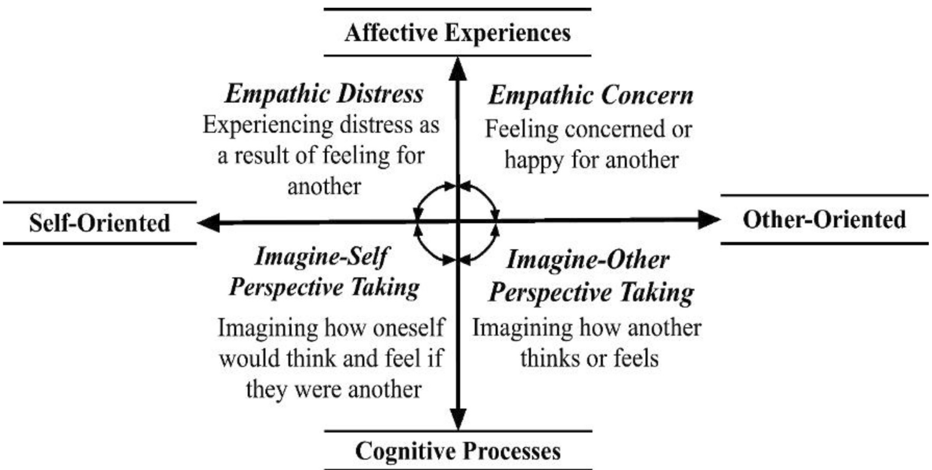

Fig. 1. Four-Part Empathy Model (Taken with permission from Hess and Fila [7]).

instrument. As indicated above, this empathy model [7] resembles the framing undergirding the Interpersonal Reactivity Index, or IRI [11]. However, two constructs of the IRI (**fantasy**, or imagining oneself as a character in books movies or plays; **personal distress**, or the tendency to become tense or anxious, in general) do not map directly onto this empathy model. Moreover, we leveraged prior work that suggested unique uses of empathy during different stages of the engineering design process [15, 16] or aspects of engineering work [1, 5]. In synthesizing these findings, we divided the survey into three broad design phases: (1) needfinding, (2) concept generation, and (3) evaluation. Next, the authors began designing potential items aligned with these empathy types.

Table 1 details the 20-item survey that resulted from our conversations. As Table 1 indicates, eight items aligned with **Imagine Self Perspective-Taking** (ISPT), seven items aligned with **Imagine-Other Perspective-Taking** (IOPT), and five items aligned with **Affective Empathy** (AE). The AE construct included both empathic distress and empathic concern elements, and thus does not distinguish between self/other orientations. By design phase, needfinding included six items, concept generation included six items, and evaluation included eight items.

4. Research Methods

4.1 Data Collection

We administered the survey to students enrolled in one of five sections of a First-Year Engineering course at a large public midwestern university at the end of the Fall 2019 semester. Students received up to 1% extra credit for participating in this study. The survey included approximately 60 items, including the 20 items included Table 1. The other 40 items did not focus on how empathy manifests in design and were not utilized in this study. Table 2 identifies the demographics of the 419 students who completed the survey. As a maximum of 600 students may have completed the survey, the response rate was roughly 70% (note, since not all sections were full, the rate is slightly higher).

4.2 Data Analysis: Confirmatory Factor Analysis (CFA)

We utilized confirmatory factor analysis (CFA) to identify the structural validity of two distinct construct configurations or measurement model types. The first set of configurations did not account for design phase, whereas the second did. Confirmatory Factor Analysis (CFA) is a type of structural equation modelling [17] wherein the objective is to test the structural validity of a measurement model. In short, CFA approaches the question, "Is the internal structure of the instrument congruent with the structure of the construct domain?" [18]. A measurement model consists of links between item responses (i.e., observed variables), latent variables (i.e., factors or constructs), and error terms. CFA assumes that each observed or measured variable (i.e., an individual's response to a survey question)

Table 1. Survey Measuring Empathy in Engineering Design

Phase	Type/No.	Subhead
Needfinding Preface: *While reading or hearing about the design scenario:*	IOPT_01	I imagined the users' everyday activities within their real-life context.
	AE_01	I felt sorry for the user experiencing the problem.
	ISPT_01	I imagined how I would feel if I experienced the problem.
	ISPT_02	I felt that I was able to relate to the challenges the users experience in their everyday life.
	ISPT_03	I imagined challenges that I would experience everyday if I were the user.
	IOPT_02	I imagined how the users would feel when they experience the problem.
Concept Generation Preface: *While generating my design ideas:*	IOPT_03	I imagined what design criteria would be the most important to the users.
	AE_02	I felt happy when generating ideas that can be helpful to the users.
	IOPT_04	I imagined how my ideas would look from the users' perspectives.
	ISPT_04	To generate more design ideas, I imagined how I would feel if I were the user.
	ISPT_05	I generated ideas by imagining that I were a user.
	AE_03	I hoped that my ideas would be useful for the users.
Evaluation Preface: *While evaluating my ideas:*	AE_04	I felt concerned when my ideas did not meet the needs of the users.
	ISPT_06	I imagined how I would use my ideas if I were the user.
	IOPT_05	I imagined why the users would like my ideas.
	IOPT_06	I imagined why the users would dislike my ideas.
	AE_05	I felt happy when my ideas helped the users.
	ISPT_07	I imagined what problems I would have when using my ideas if I were the user.
	IOPT_07	I imagined what aspects of my ideas that users would find enjoyable.
	ISPT_08	I evaluated my ideas by imagining that I were the user.

* Responses were on 7-point Likert-type scale where 1 = Not at all true of me and 7 = Very true of me.

Table 2. Participant Overview

Description	N
Total Participants	419
Academic Standing	
Freshman	390
Sophomore	25
Junior	3
Not Declared or Unknown	1
Sex	
Male	312
Female	104
Other or Decline to Specify	3
Race/Ethnicity	
American Indiana or Alaska Native	2
Asian	117
Black or African American	12
Hispanic or Latino	28
Native Hawaiian or Pacific Islander	2
White or Caucasian	277
Other	2
Not Declared	10
Age (M, SD)	18.4 (0.60)

has two causes: the factor (or latent variable) and an error term. In the measurement model, the relationships between latent and observed variables are like regression statistics.

4.2.1 Evaluation of Assumptions for CFA

We utilized MacCallum, Brown, and Sugawara [19] as a guide for sample size requirements based on degrees of freedom when seeking a statistical power of .80. In their estimation, a "close fit" model with 20 degrees of freedom requires 435 participants, whereas a close fit model with 10 degrees of freedom requires 782 participants. Thus, models that are less complex may have accompanying statistical power concerns when sample sizes are low.

Data tended to be non-normal. To offset these concerns, we computed Santora-Bentley (SB) modified statistics for major statistical indices. Next, we checked data pairings for linearity by producing and reviewing scatterplots for a subset of random pairings. While most pairings appeared linear, combinations with one item were potentially problematic (AE_04). While we recognize this potential limitation, we did not modify this variable. Finally, we had no missing data, as participants were required to respond to all items.

4.2.2 Fit Indices

The first and primary test of interest in our CFA was the chi-square test. We inferred good model fit if we failed to reject the null hypothesis, although, this inference is not exactly what this test ascertains. As MacCallum et al. [19] state:

> If the null hypothesis is not rejected, we conclude that the data are not sufficiently inconsistent with the null hypothesis for us to reject that hypothesis. This latter outcome does not imply clear support for the model but rather the absence of strong evidence against it. (p. 135).

Second, we sought thresholds identified by Schreiber et al. [20], including a SB-modified Tucker-Lewis Index (TLI) greater than 0.95 and an SB-modified root mean squared error approximation statistic less than 0.08 (ideally, below 0.06).

Finally, if a measurement model was rejected based on the chi-square test, we more closely scrutinized variables and tested potential interdependence (i.e., correlation) of error terms. This re-evaluation led to the generation and testing of new measurement models and, in turn, a re-examination of the chi-squared statistic in any new model as compared to its predecessor (along with the other fit indices). As Schreiber et al. [20] stated, "If a model has been modified and reanalyzed, one should provide evidence that the modified model is statistically superior to the original model with a chi-square test" (p. 327).

5. Results

5.1 Descriptive Statistics and Inter-Item Correlations

This section provides a descriptive overview of students' responses to the empathy survey items. This serves as context for the subsequent sections and provides a sense of students' self-described level of empathy. In addition, prior to conducting factor analyses, it is warranted to examine the correlations between items. Due to space considerations, we only present findings of inter-item correlations by empathy type (rather than across empathy types).

Tables 3, 4, and 5 provide descriptive statistics for ISPT, IOPT, and AE, respectively. All item responses aligned with a seven-point Likert-type scale, with seven indicating strong agreement and one indicating strong disagreement. For the purposes of instrument design, our primary interest in these Tables were the inter-item correlations which indicate potential alignment for CFA. All inter-item correlations were significant ($p < 0.05$). In addition, we utilized Cohen's [21] criterion to identify effect sizes, wherein $r > 0.50$ suggests a large effect, $r > 0.30$ suggests a medium effect, and $r > 0.10$ suggests a small effect.

ISPT included 28 inter-item correlations: six exhibited large effect sizes, 21 exhibited medium effect sizes, and one exhibited a small effect size (ISPT_02 and ISPT_08, $r = 0.27$). IOPT included 21

Table 3. Imagine-Self Perspective-Taking (ISPT) Descriptive Statistics and Inter-Item Correlations

Item	***M***	***SD***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
ISPT_01	5.22	1.39	1	0.60	0.63	0.40	0.38	0.44	0.40	0.36
ISPT_02	5.15	1.42		1	0.65	0.37	0.35	0.33	0.34	0.27
ISPT_03	5.36	1.33			1	0.41	0.47	0.45	0.43	0.44
ISPT_04	5.61	1.19				1	0.64	0.50	0.50	0.50
ISPT_05	5.63	1.15					1	0.48	0.52	0.58
ISPT_06	5.66	1.10						1	0.61	0.59
ISPT_07	5.78	1.07							1	0.60
ISPT_08	5.70	1.18								1

Table 4. Imagine-Other Perspective-Taking (IOPT) Descriptive Statistics and Inter-Item Correlations

Item	***M***	***SD***	**1**	**2**	**3**	**4**	**5**	**6**	**7**
IOPT_01	5.62	1.14	1	0.38	0.43	0.47	0.47	0.41	0.39
IOPT_02	5.33	1.29		1	0.31	0.43	0.46	0.39	0.39
IOPT_03	5.95	1.11			1	0.66	0.51	0.44	0.48
IOPT_04	5.76	1.06				1	0.56	0.49	0.54
IOPT_05	5.60	1.13					1	0.60	0.65
IOPT_06	5.54	1.21						1	0.46
IOPT_07	5.53	1.22							1

Table 5. Affective Empathy (AE) Descriptive Statistics and Inter-Item Correlations

Item	M	***SD***	**1**	**2**	**3**	**4**	**5**
AE_01	4.20	1.63	1	0.35	0.18	0.36	0.34
AE_02	5.61	1.25		1	0.53	0.36	0.55
AE_03	6.14	1.02			1	0.36	0.53
AE_04	5.32	1.36				1	0.44
AE_05	5.71	1.22					1

inter-item correlations: six exhibited large effect sizes, 15 exhibited moderate effect sizes, and no inter-item correlations exhibited small effect sizes. Finally, AE included 15 inter-item correlations: three exhibited large effect sizes, six exhibited moderate effect sizes, and one exhibited a small effect size (AE_01 and AE_03, r = 0.18). Overall, given the abundant large/medium and few small effect sizes, inter-item correlations suggested factorability of the data.

5.2 Confirmatory Factor Analysis: Round 1

This section details factor structures derived from seeking to validate the empathy constructs by empathy type alone. Thus, in "Round 1" we do not account for potential variation by design phase.

5.2.1 Imagine Self Perspective-Taking (ISPT)

We initially generated a CFA model estimating ISPT that included all eight potential items (see Table 1). The model fit was unacceptable, χ^2 (20) = 201.70, p < 0.01, RMSEA = 0.147, 90% CI [0.169, 0.206], TLI = 0.771. Hence, we removed items with the smallest factor loading stepwise until the model fit the specified objectives. First, we removed ISPT_02, and the model improved but remained unacceptable, χ^2 (14) = 95.43, p < 0.01, RMSEA = 0.118, 90% CI [0.134, 0.178], TLI = 0.835. Second, we removed ISPT_01. Again, the model improved but remained unacceptable, χ^2 (9) = 38.57, p < 0.01, RMSEA = 0.089, 90% CI [0.086, 0.142], TLI = 0.937. Third, we removed ISPT_01, The χ^2 value improved but the model remained unacceptable, χ^2 (5) = 31.82, p < 0.01, RMSEA = 0.113, 90% CI [0.118, 0.191], TLI = 0.912. Finally, we removed ISPT_04 and achieved an acceptable model fit, χ^2 (2) = 5.90, p = 0.052, RMSEA < 0.068, 90% CI [0.031, 0.151], TLI = 0.976. Since the RMSEA was still above 0.06, we sought to strengthen the model by correlating error terms between ISPT_05 and ISPT_08. Afterwards, the model notably improved: χ^2 (1) = 0.68, p = 0.41, RMSEA < 0.01, 90% CI [0.000, 0.127], TLI = 1.004. Fig. 2 shows a graphical depiction of the measurement model estimating the accepted solution for ISPT.

5.2.2 Imagine-Other Perspective-Taking (IOPT)

Next, we generated a CFA model estimating IOPT that included all seven potential items (see Table 1). The model fit was unacceptable, χ^2 (14) = 45.43, p < 0.01, RMSEA = 0.073, 90% CI [0.084, 0.129], TLI = 0.916. As the RMSEA was below 0.08, thus fitting one of the fit indices, rather than removing any items, we correlated error terms between IOPT_03 and IOPT_04. This generated an acceptable model fit, χ^2 (13) = 0.67, p = 0.13, RMSEA = 0.033, 90% CI [0.033, 0.085], TLI = 0.988. Fig. 3 shows the accepted measurement model estimating IOPT. Note, this was the only measurement model tested

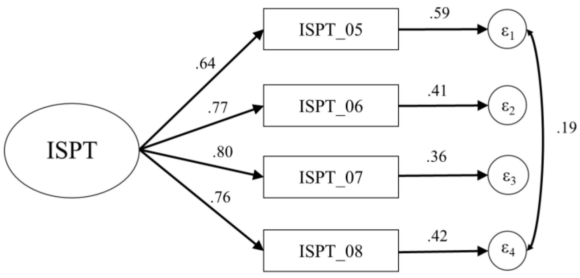

Fig. 2. CFA Measurement Model Estimating Imagine-Self Perspective-Taking (ISPT).

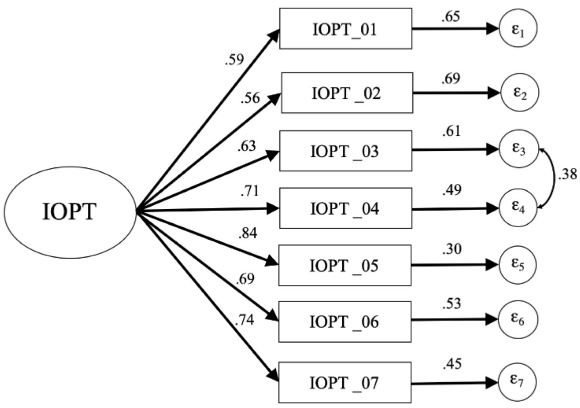

Fig. 3. CFA Measurement Model Estimating Imagine-Other Perspective-Taking (IOPT).

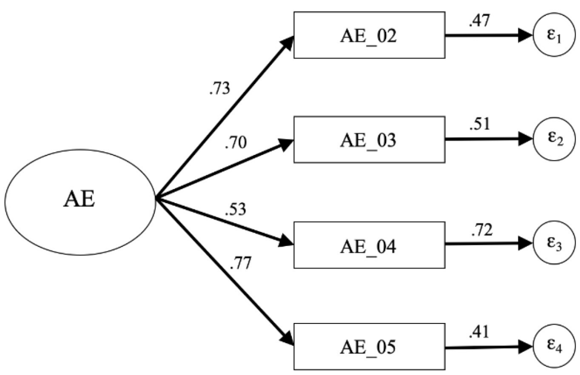

Fig. 4. CFA Model Estimating Affective Empathy (AE).

in Round 1 that did not require the removal of any items.

5.2.3 Affective Empathy (AE).

We next generated a measurement model estimating AE that included all five potential items (see Table 1). The model fit was unacceptable, χ^2 (5) = 21.30, p < 0.01, RMSEA = 0.088, 90% CI [0.073, 0.148], TLI = 0.918. Before removing any additional items, we attempted to correlate error terms, but we were unable to find a robust solution. Thus, we removed AE_01, as it had the smallest loading. As a result of this exclusion, the model fit was found to be acceptable, χ^2 (2) = 2.00, p = 0.367, RMSEA = 0.002, 90% CI [0.000, 0.116], TLI = 1.00.

5.3 Confirmatory Factor Analysis: Round 2

While the initial measurement models theorized that we did not need to distinguish between potential variation across design phases, we noted that modifications to ISPT and AE both involved the removal of items at or towards the front end of the design phases (i.e., Needfinding, Concept Generation). Further, the substantial removal of items on ISPT introduced potential concerns regarding both Type I and Type II errors (i.e., less complex models require larger sample sizes).

Thus, we reconsidered the configuration of measurement models. Given the design of the instrument also included three design phases representing three distinct sections of the survey (see Table 1), we posited that a more valid structure might account for variation across the design phases. Hence, we tested novel measurement models via CFA, but here positing that each empathy type manifests across design phases in a distinct way and is thus distinctly measurable (i.e., there are distinct subconstructs that should be tested that account for design phase). This section presents these revised CFA measurement models.

5.3.1 Imagine Self Perspective-Taking (ISPT)

We generated a CFA model estimating ISPT that included all eight potential items, but here we distinguished by design phase, wherein three items loaded onto needfinding, two items onto concept generation, and three items on evaluation (see Table 1). We theorized that these three latent variables would each be correlated with one another, thus introducing three correlation paths. The model fit was good but not quite acceptable, χ^2 (17) = 30.56, p = 0.022, RMSEA = 0.044, 90% CI [0.040, 0.084], TLI = 0.972. Hence, we began testing models with correlated error terms. After multiple tests, we found an acceptable model fit by correlating error terms between ISPT_05 and ISPT_08, χ^2 (16) = 22.85, p = 0.12, RMSEA = 0.032, 90% CI [0.024, 0.074], TLI = 0.982. Fig. 5 shows the final measurement model estimating three distinct but correlated ISPT constructs by design phase.

While the RMSEA and TLI values are less ideal than the ISPT model tested in Round 1, this model has substantively higher degrees of freedom when compared to the Round 1 model (16 versus 1). Thus, this model has more statistical power. In both models, there is a correlation between error terms. Hence, we argue that the ISPT measurement model 4 that accounts for design phase (see Fig. 4) was superior to the ISPT model that does not (see

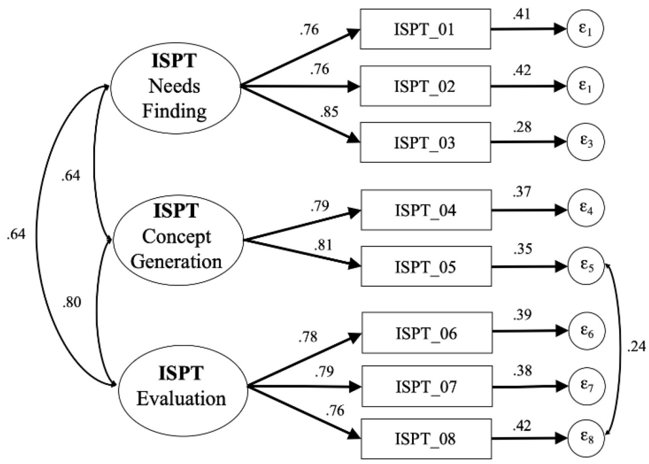

Fig. 5. CFA Measurement Model Estimating Imagine-Self Perspective-Taking (ISPT) By Design Phase.

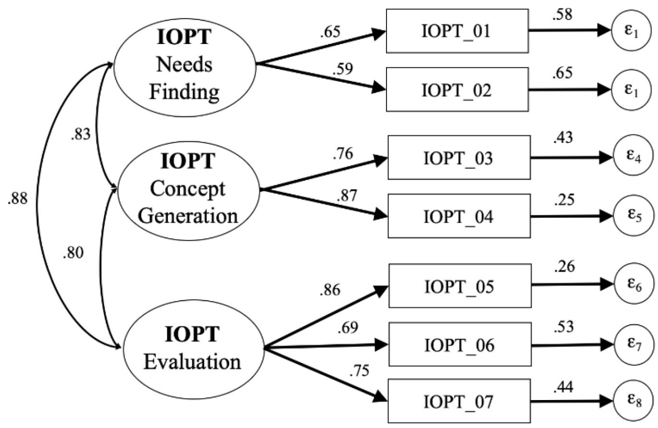

Fig. 6. CFA Measurement Model Estimating Imagine-Other Perspective-Taking (ISPT) By Design Phase.

Fig. 1). In brief, these considerations support distinguishing among empathy types by design phase.

5.3.2 Imagine-Other Perspective-Taking (IOPT)

As with ISPT, we generated a CFA model estimating IOPT that included seven potential items divided by design phase: two items aligned with needfinding, two items aligned with concept generation, and three items aligned with evaluation (see Table 1). The model fit was acceptable, χ^2 (11) = 12.86, p = 0.30, RMSEA = 0.048, 90% CI [0.015, 0.078], TLI = 0.995. Fig. 6 depicts this model.

This model has less degrees of freedom than the prior model and, therefore, less statistical power. In addition, the RMSEA is less ideal. However, the TLI improved and we did not need to correlate error terms. Thus, based on the statistics alone, it is less apparent that the IOPT measurement model that accounts for design phase is superior when compared to the two distinct ISPT configurations.

5.3.3 Affective Empathy (AE).

Finally, we generated a CFA model estimating Affective Empathy that included four of the five items. Specifically, two items aligned with concept generation and two items aligned with evaluation (see Table 1). This model was acceptable, χ^2 (1) = 0.12, p = 0.73, RMSEA < 0.01, 90% CI [0.000, 0.100], TLI = 1.019. The RMSEA and TLI statistics were superior in this model than the prior model (Fig. 4), thus supporting account for design phase when measuring AE (see Fig. 7).

5.4 Internal Consistency Reliability Check

We computed CFAs from two distinct theoretical starting points. While we were able to realize acceptable models in both instances, we note that the final solutions in the initial set of models included more items per construct than the latter. Thus, while the latter set of models may be more theoretically valid, it may also have greater concerns pertaining to internal reliability when compared to the first set of models.

To identify the extent to which each set of constructs includes an internally consistent set of items, we computed internal consistency coeffi-

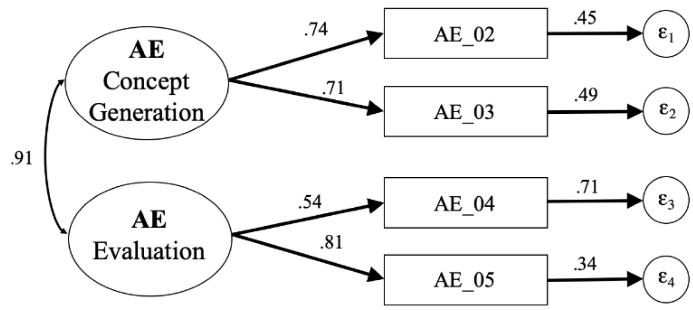

Fig. 7. CFA Measurement Model Estimating Affective Empathy (AE) By Design Phase.

Table 6. Empathy Types without Distinction by Design Phase

Construct	Items	α
ISPT	ISPT_05; ISPT_06; ISPT_07; ISPT_08	0.84
IOPT	IOPT_01; IOPT_02; IOPT_03; IOPT_04; IOPT_05; IOPT_06; IOPT_07	0.86
AE	AE_02; AE_03; AE_04; AE_05	0.77

Table 7. Empathy Types with Distinction by Design Phase

Construct	Items	α
ISPT – Needfinding	ISPT_01; ISPT_02; ISPT_03	0.83
ISPT – Concept Generation	ISPT_04; ISPT_05	0.78
ISPT – Evaluation	ISPT_06; ISPT_07; ISPT_08	0.82
IOPT – Needsfinding	IOPT_01; IOPT_02	0.55
IOPT – Concept Generation	IOPT_03; IOPT_04	0.79
IOPT – Evaluation	IOPT_05; IOPT_06; IOPT_07	0.80
AE – Concept Generation	AE_02; AE_03	0.68
AE – Evaluation	AE_04; AE_05	0.61

cients (i.e., Cronbach's alpha) for both configurations of measurement models. We utilized DeVellis [22] as a guide to ascertain the extent of internal consistency reliability. Per DeVellis' guidance, alpha statistics above 0.70 are acceptable, values above 0.60 are minimally acceptable, and any value below 0.60 is unacceptable.

Table 6 shows results for the final construct solutions from Round 1. As Table 6 shows, each of the constructs representing empathy types had acceptable internal consistency reliability.

Table 7 shows the results for the final construct solutions from Round 2. Here, most statistics were greater than 0.70, and thus acceptable. Two constructs were below 0.70 but above 0.60 (AE – Concept Generation and AE – Evaluation). Finally, one construct was unacceptable (IOPT – Needfinding). Thus, IOPT – Needsfinding was unacceptable, AE – Concept Generation and AE – Evaluation were minimally acceptable, and ISPT – Needfinding, ISPT – Concept Generation, IOPT – Concept Generation, and IOPT – Evaluation were acceptable.

These findings support the notion that, at this stage, constructs from Round 1 tend to have greater internal consistency reliability than those generated in Round 2, and thus may be more practically useful. Future work needs to design and test additional items that account for the manifestation of these empathy types across these design phases.

6. Discussion

6.1 Summary

This study utilized confirmatory factor analysis [23] to ascertain the structural validity [24] of three empathy types. We tested two distinct construct configurations, including one which did (Round 1) and one which did not (Round 2) account for how these empathy types manifest distinctly across three design phases. While accounting for design phase generally led to an inclusion of more survey items in measurement models and improved model fit, the limited number of items on some of these constructs still results in pragmatic issues regarding internal consistency reliability. Thus, at this stage, the measurement models that do not account for design phase may be more reliable, but the measurement models that distinguish by design phase may be more theoretically valid.

While some of the sub-constructs that account for design phase require additional work for future use, we can still draw inferences from these accepted models. Specifically, the measurement models show that, while empathy types manifest distinctly by design phase, the distinct constructs remain strongly correlated. For example, while Imagine-Other Perspective-Taking appears to manifest distinctly in needfinding, concept generation, and evaluation, there is a strong correlation between each. These same considerations are also applicable for Affective Empathy and Imagine-Self Perspective-Taking.

Notably, the two distinct measurement model configurations (i.e., accounting for and not accounting for design phase) that comprised Affective Empathy contained the same set of four items. The item that was removed in each was the single item pertaining to needfinding. Before we can make any empirical claims associated with affective empathy, we need to generate and test additional items that theoretically underlie "Affective Empathy – Needfinding." Here, revisiting the theoretical framework utilized to guide the instrument design, we also argue that explicitly differentiating self-oriented and other-oriented empathy types is important, as *empathic distress* is generally argued as integral to motivating interpersonal helping behavior [12].

6.2 Recommendations for Use

The instrument designed in this study has several potential uses in engineering design instructional contexts. First, instructors may use student survey scores to support course and project decisions. For example, instructors may emphasize empathy within design stages for which their students score poorly or assign activities that focus on under-utilized empathy types. As a hypothetical example based on the data found here, instructors might find that students tend to empathize more greatly at the outset of the project but not towards the end of the design process. Thus, instructors might identify the need for additional empathy scaffolds in later design stages.

In alignment with the above consideration, the survey can also be used to support students' reflective practice in design learning both before and after project work. The survey starts by asking the respondent to describe a recent design project, followed by a list of Likert-type items with "I tend to" statements. These statements are also situated in the design stages. By reflecting before a project, students may be primed to recognize how empathy informs their process and may reflect-for-action regarding how they will incorporate empathy into their work across design stages. By reflecting after a project, students may develop a stronger realization of their own emphatic tendencies and set goals for future design projects.

Finally, instructors might also utilize preliminary survey scores to inform team formation and to identify teams that might need specific empathy scaffolding or support. In a study of biomedical engineering students' empathic design in teams, where this survey instrument was used, Kim and colleagues [21] found notable qualitative differences between the high and low empathy teams. In this study, the team with the highest survey score interacted with a high number of and variety of stakeholders. These interactions led to incremental variations in their design and a successful project outcome, as opposed to major pivots experienced within teams who exhibited lower empathy. Thus, this finding suggests the use of the instrument for formative feedback, such as identifying potential challenges experienced by low-empathizing teams when engaging in human-centered design, as well as the need to provide additional guidance to improve these students' user interactions.

6.3 Limitations and Future Work

Both configurations tested here presented acceptable measurement models, with the measurement models accounting for design phase as more theoretically robust. This finding suggests three areas of future work that are needed to improve this instrument's design.

First, additional items need to be generated for the empathy constructs that account for design phase. In the instance of Affective Empathy, only one-item loaded onto Needfinding and we were therefore unable to generate and test a measurement model for this potential construct in this study. In addition, five of these constructs contained only two items; we would suggest at least three items per sub-construct are needed to reduce potential concerns associated with internal consistency reliability in future instrument use.

Second, we hypothesize that the Affective Empathy construct should be separated into empathic distress and empathic concern. While parsing between empathic distress and empathic concern was a challenge during item development, these are recognized as distinct phenomena [10]. For example, empathic distress was the key component of Hoffman's [12] moral developmental model. Therein, empathic distress represented internalized feelings which inspired helping behavior. Conversely, in an interview-based study, engineering students described empathic concern more as a guiding force in project selection, i.e., bringing them to project in which they had the opportunity to help people [25]. Thus, one potential way to operationally distinguish concern from distress would be to further identify and then emphasize distinct motivational aspirations or behavioral tendencies aligned with other empathy types.

Finally, we only tested configurations accounting for how empathy varies across three design phases: needfinding, concept generation, and evaluation. While the parsimony of a tripartite design phase model might be more practically useful, it also might neglect other key phases of engineering design wherein empathy manifests. Thus, another future research step involves collecting observational and interview data to see how students utilize and experience empathy in other phases or more nuanced stages of engineering design. Identification of additional or modified empathically salient design stages will require the development and testing of new items for each stage.

7. Conclusions

The objective of this study was to design and ascertain the structural validity of an engineering-specific measure of empathy that was applicable for use in engineering design instructional contexts. We conducted confirmatory factor analysis on two distinct measurement models. The first model consisted of three constructs representing distinct empathy types: imagine-other perspective-taking, imagine-self perspective tasking, and affective empathy. The second model consisted of the same three constructs, but here the constructs were divided into three sub-constructs based on the design stage (needfinding, concept generation, and evaluation) during which the empathy type manifested. The second model proved acceptable and preferable over the first model, based on the inclusion of more survey items and improved model fit. However, internal consistency reliability on some of the sub-constructs remains a concern for immediate use of this instrument. These findings suggest that different empathy types are relevant across the engineering design process and that these empathy types manifest differently across distinct design stages. We hope that this study can provide guidance for instructors who aspire to assess empathy throughout their students' design processes and can allow instructors to provide intentional formative feedback to bolster students' engineering design skills. However, based on the limitations outlined above, additional survey development is needed to account for a more holistic representation of how empathy manifests throughout engineering design.

Acknowledgments – This work is based upon work supported by the National Science Foundation under Grant DUE #1738214. Any opinions, findings, and conclusions or recommendations expressed in this paper, however, are those of the authors and do not necessarily reflect reviews of the NSF.

References

1. M. Kouprie and F. Sleeswijk Visser, A framework for empathy in design: Stepping into and out of the user's life, *Journal of Engineering Design,* **20**(5), pp. 437–448, 2009.
2. J. L. Hess, N. D. Fila and S. Purzer, The relationship between empathic and innovative tendencies among engineering students, *International Journal of Engineering Education*, **32**(3A), pp. 1236–1249, 2016.
3. J. Walther, S. E. Miller and N. W. Sochacka, A model of empathy in engineering as a core skill, practice orientation, and professional way of being, *Journal of Engineering Education,* **106**(1), pp. 123–148, 2017.
4. J. Strobel, J. L. Hess, R. C. Pan and C. A. Wachter Morris, Empathy and care within engineering: Qualitative perspectives from engineering faculty and practicing engineers, *Engineering Studies*, **5**(3), pp. 137–159, 2013.
5. C. Rasoal, H. Danielsson and T. Jungert, Empathy among students in engineering programmes, *European Journal of Engineering Education*, **37**(5), pp. 427–435, 2012.
6. N. D. Fila and J. L. Hess, In their shoes: Student perspectives on the connection between empathy and engineering, presented at the *American Society for Engineering Education Annual Conference*, New Orleans, LA, 2016.
7. J. L. Hess and N. D. Fila, The development and growth of empathy among engineering students, presented at the *American Society for Engineering Education Annual Conference*, New Orleans, LA, 2016.
8. M. A. Clark, M. M. Robertson and S. Young, "I feel your pain": A critical review of organizational research on empathy, *Journal of Organizational Behavior*, **40**(2), pp. 166–192, 2019.
9. J. L. Hess, J. Strobel and R. C. Pan, Voices from the workplace: Practitioners' perspectives on the role of empathy and care within engineering, *Engineering Studies*, **8**(3), pp. 212–242, 2016.
10. C. D. Batson, These things called empathy: Eight related but distinct phenomenon, in *The Social Neuroscience of Empathy*, J. Decety and W. Ickes Eds. Cambridge, MA: MIT Press, ch. 1, pp. 3–15, 2009.
11. M. H. Davis, Measuring individual differences in empathy: Evidence for a multidimensional approach, *Journal of Personality and Social Psychology*, **44**(1), pp. 113–126, 1983.
12. M. L. Hoffman, *Empathy and moral development: Implications for caring and justice*, Cambridge, UK: Cambridge University Press, 2000.
13. J. C. Oxley, *The moral dimensions of empathy: Limits and applications in ethical theory and practice*, New York, NY: Palgrave Macmillan, 2011.
14. S. Baron-Cohen, *The science of evil: On empathy and the origins of cruelty*, New York: Basic Books, 2011.
15. N. D. Fila, J. L. Hess, Ş. Purzer and E. Dringenberg, Engineering students' utilization of empathy during a non-immersive conceptual design task, *International Journal of Engineering Education,* **32**(3B), pp. 1336–1348, 2016.
16. J. L. Hess and N. D. Fila, The manifestation of empathy within design: Findings from a service-learning course, *CoDesign: International Journal of CoCreation in Design and the Arts*, **12**(1–2), pp. 93–111, 2016.

17. B. G. Tabachnick and L. S. Fidell, *Using multivariate statistics*, 6th ed. Harlow: Pearson, 2014.
18. K. A. Douglas, A. Rynearson, S. Purzer and J. Strobel, Reliability, validity, and fairness: A content analysis of assessment development publications in major engineering education journals, *International Journal of Engineering Education*, **32**(5(A)), pp. 1960–1971, 2016.
19. R. C. MacCallum, M. W. Browne and H. M. Sugawara, Power analysis and determination of sample size for covariance structure modeling, *Psychological Methods,* **1**(2), p. 130, 1996.
20. J. B. Schreiber, A. Nora, F. K. Stage, E. A. Barlow and J. King, Reporting structural equation modeling and confirmatory factor analysis results: A review, *The Journal of Educational Research*, **99**(6), pp. 323–338, 2006.
21. J. Cohen, A power primer, *Psychological Bulletin*, **112**(1), pp. 155–159, 1992.
22. R. F. DeVellis, *Scale development: Theory and applications*, Los Angeles, CA: SAGE Publications, Inc., 2011.
23. B. Thompson, *Exploratory and confirmatory factor analysis: Understanding concepts and applications*, Washington DC: American Psychological Association, 2004.
24. S. Messick, Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning, *American Psychologist*, **50**(9), p. 741, 1995.
25. N. D. Fila and J. L. Hess, Exploring the role of empathy in a service-learning design project, presented at the *DTRS 10: Design Thinking Research Symposium 2014*, West Lafayette, IN, 2014.

Justin L. Hess is an assistant professor in the School of Engineering Education at Purdue University. Dr. Hess received each of his degrees from Purdue University, including a PhD in Engineering Education, a Master of Science in Civil Engineering, and a Bachelor of Science in Civil Engineering. His primary research interests focus on several interrelated spaces: empathy; engineering ethics; diversity, equity, and inclusion; design; and innovation. He is currently the division chair-elect of the Liberal Education/Engineering & Society Division (LEES) of ASEE.

Nicholas D. Fila is a research assistant professor in Electrical and Computer Engineering at Iowa State University. He earned a BS in Electrical Engineering and a MS in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign and a PhD in Engineering Education from Purdue University. His current research interests include innovation, empathy, engineering design, course design.

Eunhye Kim is a PhD student and research assistant in the School of Engineering Education at Purdue University. Her research interests lie in engineering design education, engineering students' social processes (shared cognition and group emotion) in interdisciplinary design and innovation projects. She earned a BS in Electronics Engineering and an MBA in South Korea and worked as a hardware development engineer and an IT strategic planner in the industry.

Senay Purzer is a Professor of Engineering Education in the School of Engineering Education at Purdue University. Her research focuses on engineering design in college and pre-college education. Purzer received a BSE with distinction in Multidisciplinary Engineering in 2009 (Arizona State University) and a B.S. degree in Physics Education in 1999 (Hacettepe University). Her MA (2002) and PhD (2008) degrees are in Science Education from Arizona State University. She is also the chief editor of the Journal of Pre-College Engineering Education Research (J-PEER).and an editorial board member of Science Education.