

Establishing a Rubric to Assess Students' Empathy Development Using Artifacts from Biomedical Engineering Courses

Stephanie J. Lunn*, Cristi L. Bell-Huff[†], Joseph M. Le Doux[‡]

The Wallace H. Coulter Department of Biomedical Engineering

Georgia Institute of Technology, Atlanta, GA 30332, USA

Email: *stephanie.lunn@bme.gatech.edu, [†]cristi.bell-huff@bme.gatech.edu, [‡]joseph.ledoux@bme.gatech.edu

Abstract—This research full paper describes the development of a rubric to assess students' empathy through their coursework. Engineering departments are increasingly striving to not just foster technical competencies but to also encourage professional skills like empathy and an other-centered service orientation. In this work, we apply Zaki's model of empathy — which conceptualizes the construct in terms of its cognitive, affective, and action-oriented components — to guide the creation of a rubric to assess students' empathy using artifacts produced. While existing tools to measure empathy typically rely on self-reports, we instead focus on a standardized approach to external evaluation. We describe the justification for the decisions made as well as the pilot assessment conducted in February of 2022. In this pilot, we applied the rubric to students' assignments from four undergraduate biomedical engineering courses seeking to cultivate empathy at a large public institution in the Southeastern United States. These assignments include: 1) user stories from a problem-based learning class; 2) empathy maps from an introductory design class; 3) user needs assessment reports from a senior capstone experience; and 4) personal narratives submitted in a non-traditional storytelling course. Through understanding of the learning activities within our department, we hope to encourage others to explore the effects of their own curriculum on the multi-dimensional development of empathy as part of engineering students' professional formation. We propose that the rubric developed in this investigation could be a useful tool applied to artifacts from other courses looking to promote empathy and to evaluate the presence of its different components.

Index Terms—Empathy, quantitative rubric, document analysis, engineering education

I. INTRODUCTION

Multiple studies conducted in science, technology, engineering, and mathematics (STEM) fields have emphasized the value of empathy in professional development and interpersonal relations [1]. Yet, understanding how empathy can be cultivated in undergraduate engineering programs remains limited. As such, we sought to explore the impact of assignments integrated throughout the biomedical engineering (BME) curriculum to foster empathy and to develop a systematic tool to assess students' applications of empathy.

We chose to create a rubric to externally evaluate components of empathy in artifacts students produce. The term “rubric” may have varying definitions and intentions for use, such as medical diagnosis or educator evaluation [2]. For the purposes of our work, we consider a **rubric** as a tool

for the assessment of students' work based on evaluation criteria, quality definitions for these criteria at varied levels, and scoring strategies [3]. They can offer instructors and researchers guidance on students' current understanding, a way to explore internalization of concepts and deliver feedback, and an opportunity to gauge empathy beyond self-reported measures.

We wanted to establish a flexible rubric that would align with programmatic outcomes while being adaptive to the diversity of learning activities employed by instructors. We operationalized empathy in the context of Jamil Zaki's model [4], which considers sharing, thinking about, and caring about others. In the course of our investigation, we worked to answer the following research questions (RQs):

- RQ1) *How should the conceptualization of empathy be translated into a quantifiable rubric?*
- RQ2) *How do we observe the components of empathy in BME students' artifacts?*

In the work that follows, we describe the theoretical framework that shaped the rubric creation further in Section II. Then, the pertinent background is provided in Section III. Section IV details our approach to the rubric development process and the justification for the decisions made. In addition, we describe its utility and a pilot assessment conducted by two researchers. Our goal in the pilot was to test the rubric on different open-ended assignments from multiple BME courses. We present the actual rubric, results of the pilot, and a discussion on the implications in Section V. In Section VI, we address our limitations. Finally, we provide a summary of the key points of this work in Section VII. This exploratory study provides preliminary insights into how the rubric can be applied to externally evaluate the components of empathy.

II. THEORETICAL FRAMEWORK

Empathy is a complex phenomenon that can be important for social interactions [5]. It has been linked to raising intercultural awareness, enhancing communication, encouraging teamwork, and promoting ethical and social justice-oriented thinking and doing [6]–[8]. However, fostering empathy can be a nebulous process, and as others have pointed out, may be influenced by individual beliefs and experiences. Previously, Kunyk and Olson performed a concept analysis around the

nature of empathy and conceptualized it as [9, p. 218]: 1) a human trait; 2) a professional state; 3) a communication process; 4) caring; and 5) a special relationship. Stosic et al. (2022) further distinguished between traits, behavioral dispositions, and abilities, which they defined as “tasks that one can perform well due to education, training, or practice” [10, p. 9]. They noted that empathy is likely, at least partially, an ability and that researchers should differentiate between the two when considering their goal. While such definitions of the concept vary throughout the literature, we base our notion of empathy on the experiences and research of Zaki, who defines it less as a fixed trait and more as a skill that can be exercised and strengthened over time [4] (closer to what Stosic et al. refer to as an ability [10]).

Zaki’s model of empathy [4] guided our inquiry, illustrated in Figure 1. It considers empathy as “an umbrella term” which can be used to describe considering, understanding, or sharing the emotions, thoughts, or needs of others and potentially being motivated to help them [4, Appendix A]. While the components of empathy may have alternative names and may activate distinct signaling pathways and structures in the brain (e.g., mentalizing versus distinguishing between the self and others) [5], [11], [12], they tend to influence each other as well [4].

As defined by Zaki’s model, “**Sharing**” involves the emotional (or affective) facet of empathy, and may also include experience sharing and personal distress [4]. It entails taking on the moods, stress, posture, or facial expressions of another. Neurologically, this component of empathy has been linked with mirror neurons, cells that induce action based on behaviors observed in another.

“**Thinking About**” refers to the cognitive aspect of empathy, also defined by mentalizing or Theory of Mind. It describes a form of mental awareness of others’ or making sense of their perspectives. It encompasses reflecting on their lived experiences, needs, emotions, and intentions.

Meanwhile, “**Caring About**” is the action-focused form of empathy that includes empathic concern and compassion. It may arise out of the thinking about and sharing the emotions of others, to drive behaviors to lend support based on perceived need. While the term is often used interchangeably with sympathy, a concept based on having pity or sorrow for someone’s misfortune, empathy is considered the next level of making connections with the spectrum of emotions from sadness to happiness. It refers to having a deeper consciousness of the other’s perspective and emotional sharing which compels the individual to willingly take steps to “do something” about the circumstances.

III. RELATED WORK

Before we describe our approach to creating a rubric for empathy, we want to first highlight the background and value of rubrics in general (Section III-A). Then, we review different approaches taken to measure empathy in Section III-B. We apply aspects of some of these existing scales in the development of our rubric.

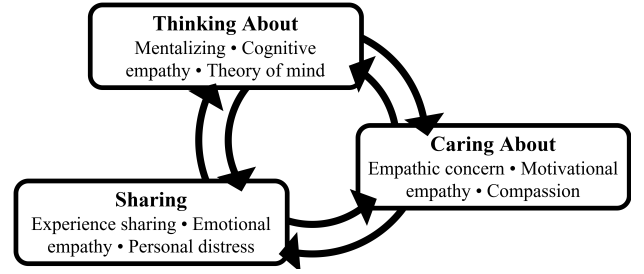


Fig. 1: Zaki’s model of empathy, adapted from [4]

A. Rubrics

Rubric development can involve many considerations. They may be more broad, to assess a particular objective or category (e.g., empathy), or specific to a particular task. Additionally, rubrics can either be analytic or holistic [13]. Analytic rubrics assess performance, characteristics, or skills on each criteria of the scales. Contrarily, an analytic rubric entails distinct assessment of multiple criteria to arrive at a whole value.

Rubrics have previously been applied in various STEM contexts for evaluation of concepts such as entrepreneurial mindset [14] and critical thinking [13], [15] in engineering, algorithm and programming in computing [16], and scientific reasoning using scientific writing of students in biology [17]. Scholars have described that rubrics can help to evaluate open-ended problems [13] and can offer an opportunity to determine effectiveness of programs, and “internal consistency among the goals, assessments, and learning experiences” [18, p. 150]. In turn, the findings from rubrics can be used to create and refine learning experiences [18]. However, it can be important to contextualize their application, and to bear in mind that assessing multi-modal items requires consideration of the academic and programmatic environment in which artifacts are assessed, as well as the scale of classes and workloads for students and teachers [19].

B. Other Tools for Measuring Empathy

Empathy is typically measured through first-person, self-appraisal scales since affective and cognitive processes may be challenging to observe directly [10]. Tools and scales to quantify empathy often arise from the fields of psychology and/or medicine. The medical applications are often in the context of physician and patient dynamics. An overview of some commonly applied scales to measure empathy based on different concepts and constructs are illustrated in Table I.

Testing empathy externally has been limited. Some assessments have applied similar notions and considered evaluation of nonverbal cues though. As an example, the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) [32] studies emotional intelligence (considering it as an ability) and defines this construct in terms of perceiving and identifying emotions, facilitating thought, understanding emotions, and managing emotions. In particular, perceiving and identifying emotions is defined as “the ability to recognize how you and those around you are feeling” [32, p. 5] and was applied in the

Empathy Questionnaire	Citation	# of Items	Scale (# of Points)	Dimensions	Focus
<i>Hogan Empathy Scale (HES)</i>	[20]	64	5	4: Social Self-Confidence, Even-Temperedness, Sensitivity, Nonconformity	Cognitive
<i>Questionnaire Measure of Emotional Empathy (QMEE) or Emotional Empathy Tendency Scale (EETS)</i>	[21]	33	9	7: Susceptibility to Emotional Contagion, Appreciation of the Feelings of Unfamiliar and Distant Others Extreme, Emotional Responsiveness, Tendency to be Moved By Others' Positive Emotional Experiences, Tendency to be Moved By Others' Negative Emotional Experiences, Sympathetic Tendency, Willingness to be in Contact with Others Who Have Problems	Affective
<i>Interpersonal Reactivity Index (IRI)</i>	[22]	28	5	4: Perspective Taking, Fantasy, Empathic Concern, Personal Distress	Cognitive and Affective
<i>Balanced Emotional Empathy Scale (BEES)</i>	[23], [24]	30	9	2: Vicarious Experience of Others' Feelings; Interpersonal Positiveness	Affective
<i>Multidimensional Emotional Empathy Scale (MDEES)</i>	[25]	30	5	6: Empathic Suffering, Positive Sharing, Responsive Crying, Emotional Attention, Feeling for Others, and Emotional Contagion	Affective
<i>Jefferson Scale of Physician Empathy (JSPE)</i>	[26]	20	7	10: Empathy, Compassion, Trust, Sympathy, Tolerance, Personal Growth, Communication, Self-Protection, Humor, Clinical Neutrality	Empathy in Physicians
<i>Empathy Quotient (EQ)</i>	[27]	60: 40 empathy, 20 "fillers"	4		Empathy in Adults
<i>Basic Empathy Scale (BES)</i>	[28]	20	5	3: Cognitive, Affective Empathy (split into Emotional Contagion and Emotional Disconnection)	Cognitive and Affective
<i>Toronto Empathy Questionnaire (TEQ)</i>	[29]	16	5	4: Emotional Contagion, Emotion Comprehension, Sympathetic Physiological Arousal, Con-Specific Altruism	Different levels of empathy
<i>Questionnaire of Cognitive and Affective Empathy (QCAE)</i>	[30]	31	4	5: Perspective Taking, Online Simulation, Emotion Contagion, Proximal Responsivity, Peripheral Responsivity	Cognitive and Affective
<i>Affective and Cognitive Measure of Empathy (ACME)</i>	[31]	36	5	3: Cognitive, Affective Empathy (split into Affective Resonance and Affective Dissonance)	Cognitive and Affective

TABLE I: Empathy questionnaires

context of exploring the cognitive process affiliated with sensing emotions observed in photographs of faces and artwork. Meanwhile, emotion management focused on behaviors and effective strategies for coping with different situations. To evaluate this component, participants were presented with social scenarios and tasked with choosing a response pertaining to different emotional outcomes. While such aspects of cognition and behavior are not explicitly labeled as empathy, they do align with Zaki's thinking and caring about others.

IV. METHODS

The initial step of rubric creation requires reflection on the broader purpose and expectations for students' performance [33]. Along these lines, our goal, as described in this work, was to develop a rubric to evaluate assignments and activities offered throughout the BME curriculum. However, existing measures for empathy often rely on internal reports rather than external evaluation. As such, we sought to create our own and to develop a general, analytic rubric to try to approach components of Zaki's empathy framework externally.

Other authors have articulated the value of naming a scale or rubric and using a descriptive title and how this can impact perceptions [10], [13]. Davis' IRI has been used as an example of how the Fantasy subscale is intended to measure the

cognitive aspect of empathy, but it tends to be more predictive of the affective components of empathy. We labeled our rubric as the External Evaluation of Empathy Rubric (EEER). In Section IV-A, we describe the learning outcomes that shaped the essential skills to be measured by this rubric. We elaborate further on the creation of the EEER in Section IV-B. In Section IV-C, we describe our pilot to evaluate the rubric. Then, we review quality considerations in Section IV-D.

A. Learning Outcomes

Well-defined learning outcomes can be important to achieve lasting change and value for students through meaningful and effective assignments, classes, courses, and programs [34]. They also require the deliberate selection of tools to measure and assess their progression. Once we defined the larger goals for the rubric, we next considered the learning outcomes pertinent to fostering empathy for students in engineering.

We surveyed the literature to explore learning activities that other educators have employed to cultivate empathy in STEM fields [1]. Next, we conducted semi-structured interviews with faculty in the BME department [35], to learn about the activities they have employed and their conceptualization of empathy and its role in engineering education. Based on the findings from the systematic literature review and our

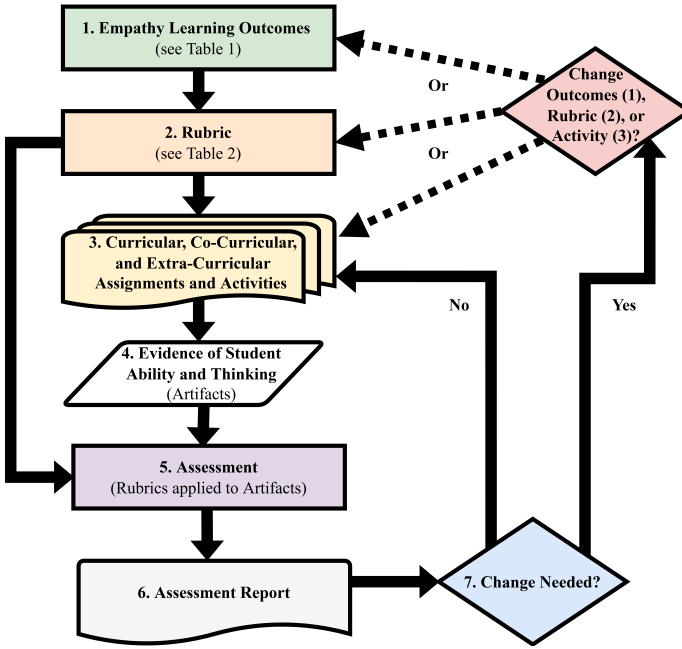


Fig. 2: Model of cultivating empathy using learning outcome driven program development and assessment, adapted from [14]

interviews, as well as Zaki’s framework, we defined a list of learning outcomes as shown in Table II. Apart from including more general objectives, such as those related to interpersonal dynamics, we also considering more engineering specific items to encompass principles around human-centered design. These outcomes are intended to not only reflect individual and/or team performance, but also to illustrate skill level and mastery for the concept of empathy and its development.

A student should be able to:

- 1) Effectively collaborate in a team setting
- 2) Demonstrate voluntary social responsibility
- 3) Apply active listening
- 4) Illustrate appropriate consideration of others’ perspectives
- 5) Establish human-centered designs
- 6) Consider broader societal implications of choices
- 7) Take action to help others
- 8) Form emotional connections with content and/or stakeholders
- 9) Display evidence of thinking about others
- 10) Articulate their internal reflection processes

TABLE II: Learning outcomes for cultivating empathy

We envision these learning outcomes being applied and assessed as described in Figure 2. Adapted from Kleine and Yoder [14], who developed this model in the context of fostering an entrepreneurial mindset, we consider the process toward the cultivation of empathy. Although the model is meant to be flexible given individual objectives, it provides a starting template for implementing, evaluating, and adapting programs and experiences.

B. Development of the External Evaluation of Empathy Rubric (EEER)

Rubric creation necessitates consideration of the overall goal as well as learning outcomes in relation to students’

tasks and expectations [13], [14]. With the learning outcomes defined, we then sought to dimensionalize them. To promote consistency and objectivity, we began an initial list using statements from existing measures described in Table I as well as Valid Assessment of Learning in Undergraduate Education (VALUE) rubrics developed by the American Association of Colleges and Universities (AAC&U). This first version included 80 items in total, taken from the following: IRI, TEQ, ACME, BES-A (Linked to Emotional Contagion and Cognitive Empathy), MDEES (Suffering and Positive Sharing), QCAE (Including items from Affective Empathy [Proximal Responsivity, Emotion Contagion, and Peripheral Responsivity] and Cognitive Empathy [Online Simulation and Perspective Taking]), EQ, and AAC&U rubrics (Civic Engagement, Intercultural Knowledge and Competence, Lifelong Learning, Problem Solving, and Ethical Reasoning).

Working together, the first two authors then met to condense and refine this list to consider how such items might fit together, align with Zaki’s framework, and be adapted from a first-person point of view. Given that we wanted a flexible tool that would be applicable across different assignments and activities, this led to consideration of the spectrum of potential aspects of empathy that students may demonstrate in their work. This revision led to the definition of five items pertaining to “Sharing,” three items pertaining to “Thinking About,” and four items pertaining to “Caring About.” The authors then further iterated over the items, refining the language and establishing parsimonious categories that could be inclusive for a range of artifacts. This resulted in the third, and final, version of the rubric, which we will present later.

C. Pilot Evaluation of Empathy Artifacts

The rubric developed was tested on artifacts collected from four courses interspersed throughout the curriculum at a large research university in the Southeastern United States from the Fall 2021 semester. These assignments included: 1) user stories from a problem-based learning (PBL) class; 2) empathy maps from an introductory design class; 3) user needs assessment reports from a senior capstone experience; and 4) personal narratives submitted in a non-traditional storytelling course. We would like to note that all artifacts examined were obtained upon receiving Institutional Review Board approval and the consent of the students.

Before evaluating the artifacts, the authors first separately reviewed the items collected, reconvening to discuss the quality levels and criteria that would inform the ratings. They negotiated on the definitions, considering what would distinguish each level of application. The pilot assessment was conducted during February of 2022, and each rater independently judged three artifacts from each of the four assignments. We will discuss what each assignment entailed further below.

User stories are informal explanations of a scenario that consider the perspective of a stakeholder. Developed as part of the inquiry phase in the engineering design process, students synthesize information gathered to foster an understanding [36]. Students were encouraged to create user stories as part of

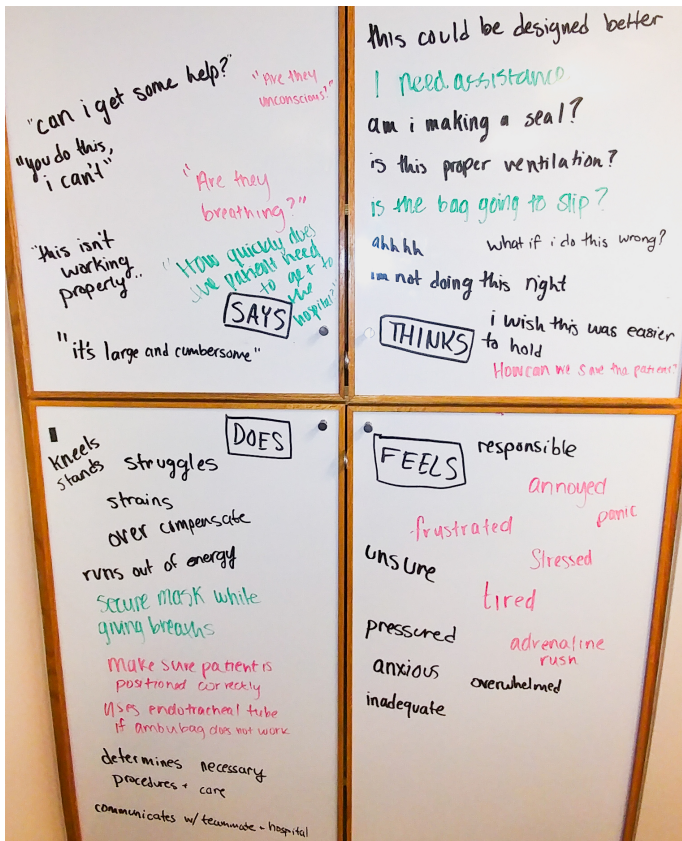


Fig. 3: Example of a team's empathy map

a PBL course in tandem with defining the problem statement, which represents the purpose of the work. An example of the problem statement and user stories created by students is illustrated in Table III.

Empathy maps are tools used to summarize learning and to elicit insights about users based on observations from research. They are visualized as four-quadrant representations to separate out what a potential user could say, think, feel, and do [37]. Figure 3 shows an empathy map developed by a group in a design class.

User needs assessments are formalized, strategic depictions and evaluations of the business and technical requirements to inform product design. These analyses are presented in terms of the design inputs, users' needs, the necessity of the feature (i.e., critical, need, or want), metric(s) needed, and reference metric(s). Frequently, teams present these as tables. We will not provide an example here as such evaluations tend to be fairly expansive to capture the full scope.

The non-traditional narratives course included prompts and the sharing of stories to promote valuable skills for personal and professional growth. Listening to narratives has also been shown to promote perspective taking [38]. One of the key features in the context of the course was to encourage students to think about their own leadership, goals, and to reflect on their experiences. An example from one of the final stories submitted is presented below, describing how a community service experience drove personal connections and fostered a

commitment to helping others:

We got to the building out of which MBG [Meals By Grace] operated early that Saturday and to my surprise there was way more people than I expected. [...] When my friends and I got to those in charge of handing out the assignments, we were told that since we were new we should do the delivery work as it would help us gain a better understanding of the community around us. [...] When we got to the first house, I got out of the car and rang the door bell not really knowing what to expect. After about 5 seconds the door swung open and out stepped the entire family of my 7th grade English teacher. Upon seeing her, I immediately went numb and very nearly dropped the package out of my hands. Here was someone who was so important in my early formative years, someone I looked up to even now, and I not once even considered that she would be having any trouble at home. She smiled and asked me how I had been, and while trembling, I answered back that I had been just fine, unsure if it was appropriate to ask the same question back, as it dawned on me that she had utilized the same smile all those years back to shield her students from the harsher realities of the world around us. We were on the clock, so I very quickly gave her my thanks for all she had done for me and walked back to the car, shook to my core about what had just happened. We continued through the rest of the drive and it became even more clear when we stopped by houses in my neighborhood and saw even more friendly but familiar faces. Now I find myself volunteering all over my community wherever and whenever I can because on that day I learned a lesson that I will carry with me forever. Help should be given to everyone around us, whether they ask [for it] or not, because we never know what is going on behind closed doors. [...] That drive and all of my other experiences within my community drove me to seek ways through which I can help those around me and that is why I finally came to the decision to pursue a career in Biomedical Engineering, through which I can combine my passions and extend my reach beyond my community and help those who cannot help themselves.

D. Quality Considerations

Construct validity is a quality measure that indicates how well a metric reflects the reality of the concept studied [39]. To establish this type of validity, we reviewed the literature (further described in [1]) and used established measures and theory to guide the development of our rubric. In addition, external validity considers how generalizable the results are [40]. In this work, we sought to develop a rubric that could be applied to different assignments and we applied it to artifacts from courses offered throughout the undergraduate

<i>Problem Statement</i>	Patients with respiratory issues (e.g. pneumonia, COVID-19) that lead to long term intubation need a more effective device to avoid excess damage to tracheal and/or smooth tissue.
<i>User Story 1</i>	A woman was brought into the ER as a victim of a multi-car crash, and despite carrying a baby she was not prioritized for the limited number of sterilized endotracheal tubes available for crash victims. There was an older man with respiratory issues who received a tube before her, despite her doctors informing her that delaying intubation could have severe damages to the fetus. Eventually she did receive a tube, but it concerning that damage could have been done to the child simply because they did not have enough tubes on hand. If the woman would have been intubated sooner it would've caused for less chance of harm to the child.
<i>User Story 2</i>	The other man was a victim of COVID-19 at the time of the car crash, and had to be intubated for long-term (>10 days). His resident (who intubated him) did not have access to a cuff manometer, as they are expensive and in limited supply - as well as taken up by the other victims of the car crash. As a result, his cuff was overinflated and caused significant damage to his smooth muscle and tracheal stenosis. Moreover, when his cuff began to deflate naturally, the 'standard' pressure range was too high for his already damaged smooth muscle. The physician later noted that they did not have any research basis for the standard pressure range, it is simply what is given to all trainees.

TABLE III: Example of students' problem statement and user stories

biomedical engineering curriculum in a pilot assessment. The authors established the definitions based on their review of the materials, discussing potential discrepancies in interpretation that could arise. For the purposes of establishing inter-rater reliability, the two authors tested the rubric separately.

V. RESULTS AND DISCUSSION

In the section that follows, we describe the EEER and the initial validation performed. Section V-A, along with the process articulated in the methods, answers RQ1 about how the conceptualization of empathy can be translated into a quantifiable rubric. The outcomes of the pilot are used to present our interpretation of RQ2 and how different artifacts may influence the aspects of empathy observed (Section V-B).

A. EEER

The final version of the EEER is illustrated in Table IV. The components of empathy are considered uniquely to allow a distinctive understanding of the presence of each. In addition, we included dimensions (or attributes) to characterize the more detailed aspects of each based on existing measures. We consider four quality performance levels — Not Evident, Somewhat Evident, Evident, and Strongly Evident — and we include the definitions for each within the rubric. We want to caution that our goal was not to make judgements about which form of empathy is needed or superior, or to claim this exists as an absolute measure that an individual may have empathy, as others have warned against [10]. Instead, we intend for the EEER to provide researchers with the opportunity to recognize the components that may be present in different artifacts and to weight them as appropriate for their intended purposes.

B. Pilot Outcomes

Assessment is important for making ongoing improvements. Evaluating the artifacts independently allowed the authors to test for differences in their interpretation. We observed all three components of empathy in the assignments examined. We present the inter-rater agreement for each facet of empathy ("Thinking About," "Sharing," and "Caring") and the overall alignment of the raters using the rubric in Table V.

With a 4-point scale, there is the possibility of selecting the same value by chance, and thus we applied Cohen's kappa to measure our agreement. In particular, we applied a linear weighted Cohen's kappa ($\hat{\kappa}_w$) since it is intended for ordinal scales where a difference of one scale point between raters is

less dramatic than variations of two points. We interpret the magnitude according to the definition in which values greater than 0.75 denotes an "excellent agreement" [46, p. 609]. We consider the $\hat{\kappa}_w$ for each component and the overall rubric alignment in Table VI. As illustrated, all are above the 0.75 threshold, individually and cumulatively.

When considering each component of empathy in terms of those where both raters agreed on the rating, the presence of each is as follows:

- **Sharing:** 21 total in agreement, construct not evident in 85.71% and present in 14.29%
- **Caring About:** 28 total in agreement, construct not evident in 42.86% and present in 57.14%
- **Thinking About:** 27 total in agreement, construct not evident in 44.44% and present in 55.56%

These breakdowns include any instance of "Somewhat Evident," "Evident," or "Strongly Evident" as present, and those ratings of "Not Evident" as such. Despite displaying them here as a variable that represents whether it exists or not, we are not trying to say an individual's empathy may exist with the same dichotomy (we perceive it as a continuum). Our intention is merely to demonstrate that the EEER can differentially recognize the disparate facets of empathy.

While cognitive and affective aspects of empathy existed in the artifacts, the construct of "Caring About" was the most prevalent facet observed in the pilot (present in 57.14%). This finding aligns with literature mentioning that behaviors are easier to observe than mental [11] and emotional [12] processes. We acknowledge that, at least partially, the presence or absence of each facet may be due to the nature of the artifacts assessed. Our pilot primarily evaluated team submissions (with the exception of the personal narratives), which could make it harder to detect individual thoughts and feelings.

Campbell and Fiske [47] have suggested that constructs should be continually validated and relationships between concepts should be assessed through multiple methods. Accordingly, we encourage others to expand upon our pilot and consider using the rubric presented for different kinds of assignments, perhaps factoring in the impact of individual versus team submissions. We also recommend evaluating artifacts using the EEER in parallel with self-reported scales to assess the convergence and divergence of each.

Empathy Component	Dimension	Strongly Evident	Evident	Somewhat Evident	Not Evident	Source References
		Displays or articulates this and/or considers the implications clearly. Provides multiple applicable examples and/or strong evidence.	Displays or articulates this and/or considers the implications sufficiently. Provides some applicable examples and/or evidence but may also provide some that are irrelevant.	May display or mention this and/or consider the implications rudimentarily. No examples or evidence provided.	Fails to display or mention this and does not consider the implications. No examples or evidence provided.	
<i>Sharing (Emotional/Affective)</i>	Displays feeling another's emotional state or their mood being altered by others (Individual or group level)					[22]–[25], [27]–[31]
	Displays their feelings or emotions about social injustice or cultural variations they observe (Societal level)					[22], [29], [41], [42]
<i>Caring About (Taking Action)</i>	Displays taking action, comforting, or putting effort into their social actions, relationships, and/or activities with others (Individual or group level)					[27], [29], [31]
	Displays taking action or community engagement to bring about change (Societal level)					[41], [42]
	Displays a solution or creating a potential solution that considers the needs of others and their worldviews (Can be any level)					[42], [43]
<i>Thinking About (Cognitive)</i>	Displays taking stock or reflecting on their own feelings, knowledge, beliefs, values, and perspectives as they relate to others' (Personal level)					[41], [44], [45]
	Displays understanding others' feelings, knowledge, beliefs, values, and perspectives (Individual or group level)					[22], [27], [28], [30], [31], [45]
	Displays understanding or consideration of community impacts, civic consequences, and/or social justice factors from different perspectives (Societal level)					[41], [42]

TABLE IV: External Evaluation of Empathy Rubric (EEER)

Sharing (Emotional/ Affective)		Rater #2				
		Strongly Evident	Evident	Somewhat Evident	Not Evident	Total
Rater #1	Strongly Evident	0	0	0	0	0
	Evident	0	3	0	0	3
	Somewhat Evident	0	0	0	3	3
	Not Evident	0	0	0	18	18
	Total	0	3	0	21	24

Caring About (Taking Action)		Rater #2				
		Strongly Evident	Evident	Somewhat Evident	Not Evident	Total
Rater #1	Strongly Evident	11	0	0	0	11
	Evident	0	3	1	1	5
	Somewhat Evident	0	0	2	6	8
	Not Evident	0	0	0	12	12
	Total	11	3	3	19	36

Thinking About (Cognitive)		Rater #2				
		Strongly Evident	Evident	Somewhat Evident	Not Evident	Total
Rater #1	Strongly Evident	7	1	0	0	8
	Evident	2	7	0	2	11
	Somewhat Evident	0	0	1	4	5
	Not Evident	0	0	0	12	12
	Total	9	8	1	18	36

Overall Rating Alignment		Rater #2				
		Strongly Evident	Evident	Somewhat Evident	Not Evident	Total
Rater #1	Strongly Evident	18	1	0	0	19
	Evident	2	13	1	3	19
	Somewhat Evident	0	0	3	13	16
	Not Evident	0	0	0	42	42
	Total	20	14	4	58	96

TABLE V: Pilot inter-rater agreement

	Weighted Kappa (κ_w)	Standard Error	Confidence Interval	
			Lower Limit	Upper Limit
Sharing (Emotional/Affective)	0.765	0.142	0.486	1
Caring About (Taking Action)	0.824	0.060	0.706	0.942
Thinking About (Cognitive)	0.776	0.070	0.638	0.914
Overall Alignment	0.817	0.040	0.738	0.896

TABLE VI: Inter-rater kappa with linear weighting

VI. LIMITATIONS

This work presents the initial steps taken to create and validate a rubric based on existing empathy measures. However, there are several limitations which should be discussed. Only two researchers were involved in developing the rubric and assessing artifacts for the validation of the EEER. Jonsson and Svingby have suggested “two raters are, under restrained conditions, enough to produce acceptable levels of inter-rater agreement” [48, p. 136]. Yet, increasing the number of individuals involved in both phases could enhance diversity in perspectives and the robustness of the reliability. In addition, we only examined a limited quantity of artifacts. Applying

the rubric more broadly (in terms of the amount assessed and using disparate artifacts from different instructors) will serve to understand its extensibility across assignments and can serve to further test its validity and reliability, as well as the variety of ways it can be interpreted. While we do not presume that data collected over the course of one semester from a single institution provides a comprehensive overview of all possible applications of our empathy rubric, the work provides insight into potential applications across a range of artifacts in BME.

We also acknowledge that the components of empathy examined can be an oversimplification. What constitutes empathy for an individual, and how it is expressed, is likely to fluctuate in magnitude and direction. Our rubric seeks to begin capturing aspects, but we encourage researchers that employ it to bear in mind the potential variations in ways students may consider empathy. We recommend proper scaffolding in lessons to promote internalization of the concept.

For instructors that include empathy in coursework or assignments, we would further like to note that it can be challenging to assess authentic empathy versus deliberate (and potentially disingenuous) attempts to include language to “earn points.” Particularly in meritocratic disciplines like engineering, this is something that could impact grading. While this is beyond the scope of what we can examine with a rubric, it is an important consideration for teaching empathy and ethics.

VII. CONCLUSIONS

There are many approaches that can be employed to cultivate empathy in students. How it can be measured is less clear. Our work describes the design of a rubric to externally evaluate empathy in BME students. The creation of the EEER was operationalized using Zaki’s model of empathy and informed by existing measures previously validated to self-assess empathy. Our pilot demonstrated that the rubric can be applied to a range of artifacts and can be used to better understand which aspects of empathy may be present. Further studies are suggested to expand upon our findings and explore empathy development in greater depth as well as in other disciplines. Future applications of the rubric could not only help to determine students’ empathy levels through coursework but could also offer insight for items produced in informal settings. Going forward, we encourage educators to continue to work towards fostering empathy in students and to contemplate the different ways empathy may manifest.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation (NSF), Award # 2025080. Any findings, conclusions, and recommendations expressed in this work do not necessarily reflect the views of the NSF.

REFERENCES

- [1] S. J. Lunn, C. L. Bell-Huff, and J. M. Le Doux, “Cultivating Inclusivity: A Systematic Literature Review on Developing Empathy for Students in STEM Fields,” in *Collaborative Network for Engineering and Computing Diversity (CoNECD)*, February 2022.

- [2] P. Dawson, "Assessment rubrics: towards clearer and more replicable design, research and practice," *Assessment & Evaluation in Higher Education*, vol. 42, no. 3, pp. 347–360, 2017.
- [3] W. J. Popham, "What's Wrong—and What's Right—with Rubrics," *Educational Leadership*, vol. 55, no. 2, pp. 72–75, 1997.
- [4] J. Zaki, *The War for Kindness: Building Empathy in a Fractured World*. Crown, 2019.
- [5] S. Héту, V. Taschereau-Dumouchel, and P. L. Jackson, "Stimulating the brain to study social interactions and empathy," *Brain Stimulation*, vol. 5, no. 2, pp. 95–102, 2012.
- [6] M. J. Riemer and D. E. Jansen, "Non-verbal intercultural communication awareness for the modern engineer," *World Transactions on Engineering and Technology Education*, vol. 2, no. 3, pp. 373–378, 2003.
- [7] K. E. Gerdes, C. A. Lietz, and E. A. Segal, "Measuring Empathy in the 21st Century: Development of an Empathy Index Rooted in Social Cognitive Neuroscience and Social Justice," *Social Work Research*, vol. 35, no. 2, pp. 83–93, 2011.
- [8] N. D. Fila and J. L. Hess, "In Their Shoes: Student Perspectives on the Connection between Empathy and Engineering," American Society for Engineering Education (ASEE) Annual Conference & Exposition, 2016.
- [9] D. Kunyk and J. K. Olson, "Clarification of conceptualizations of empathy," *Journal of Advanced Nursing*, vol. 35, no. 3, pp. 317–325, 2001.
- [10] M. D. Stosic, A. A. Fultz, J. A. Brown, and F. J. Bernieri, "What is your empathy scale not measuring? The convergent, discriminant, and predictive validity of five empathy scales," *The Journal of Social Psychology*, vol. 162, no. 1, pp. 7–25, 2022.
- [11] R. P. Spunt, A. B. Satpute, and M. D. Lieberman, "Identifying the What, Why, and How of an Observed Action: An fMRI Study of Mentalizing and Mechanizing during Action Observation," *Journal of Cognitive Neuroscience*, vol. 23, no. 1, pp. 63–74, 2011.
- [12] R. P. Spunt and M. D. Lieberman, "An integrative model of the neural systems supporting the comprehension of observed emotional behavior," *Neuroimage*, vol. 59, no. 3, pp. 3050–3059, 2012.
- [13] P. Ralston and C. Bays, "Refining a Critical Thinking Rubric for Engineering," in *ASEE Annual Conference & Exposition*, 2010, pp. 1–16.
- [14] R. E. Kleiner and J.-D. Yoder, "Operationalizing and Assessing the Entrepreneurial Mindset: A Rubric Based Approach," *The Journal of Engineering Entrepreneurship*, vol. 2, no. 2, pp. 61–90, 2011.
- [15] K. Alfrey and E. Cooney, "Developing a Rubric to Assess Critical Thinking in Assignments with an Open Ended Component," in *ASEE Annual Conference & Exposition*, 2009, pp. 14–444.
- [16] N. d. C. Alves, C. G. von Wangenheim, J. C. R. Hauck, and A. F. Borgatto, "A Large-scale Evaluation of a Rubric for the Automatic Assessment of Algorithms and Programming Concepts," in *51st ACM Technical Symposium on Computer Science Education (SIGCSE'20)*. ACM, 2020, pp. 556–562.
- [17] B. E. C. Timmerman, D. C. Strickland, R. L. Johnson, and J. R. Payne, "Development of a 'universal' rubric for assessing undergraduates' scientific reasoning skills using scientific writing," *Assessment & Evaluation in Higher Education*, vol. 36, no. 5, pp. 509–547, 2011.
- [18] K. Shively, K. M. Stith, and L. D. Rubenstein, "Measuring What Matters: Assessing Creativity, Critical Thinking, and the Design Process," *Gifted Child Today*, vol. 41, no. 3, pp. 149–158, 2018.
- [19] R. E. Burnett, A. Frazee, K. Hanggi, and A. Madden, "A Programmatic Ecology of Assessment: Using a Common Rubric to Evaluate Multimodal Processes and Artifacts," *Computers and Composition*, vol. 31, pp. 53–66, 2014.
- [20] R. Hogan, "Development of an Empathy Scale," *Journal of Consulting and Clinical Psychology*, vol. 33, no. 3, p. 307, 1969.
- [21] A. Mehrabian and N. Epstein, "A measure of emotional empathy," *Journal of Personality*, vol. 40, pp. 525–543, 1972.
- [22] M. H. Davis, "Measuring Individual Differences in Empathy: Evidence for a Multidimensional Approach," *Journal of Personality and Social Psychology*, vol. 44, no. 1, pp. 113–126, 1983.
- [23] A. Mehrabian, "Manual for the Balanced Emotional Empathy Scale (BEES)," 1996, Unpublished Manuscript.
- [24] A. Mehrabian, "Relations among personality scales of aggression, violence, and empathy: Validation evidence bearing on the Risk of Eruptive Violence Scale," *Aggressive Behavior*, vol. 23, no. 6, pp. 433–445, 1997.
- [25] D. R. Caruso and J. D. Mayer, "A Measure of Emotional Empathy for Adolescents and Adults," 1998, Unpublished Manuscript.
- [26] M. Hojat, J. S. Gonnella, T. J. Nasca, S. Mangione, M. Vergare, and M. Magee, "Physician Empathy: Definition, Components, Measurement, and Relationship to Gender and Specialty," *American Journal of Psychiatry*, vol. 159, no. 9, pp. 1563–1569, 2002.
- [27] S. Baron-Cohen and S. Wheelwright, "The Empathy Quotient: An Investigation of Adults with Asperger Syndrome or High Functioning Autism, and Normal Sex Differences," *Journal of Autism and Developmental Disorders*, vol. 34, no. 2, pp. 163–175, 2004.
- [28] D. Jolliffe and D. P. Farrington, "Development and validation of the Basic Empathy Scale," *Journal of Adolescence*, vol. 29, no. 4, pp. 589–611, 2006.
- [29] R. N. Spreng*, M. C. McKinnon*, R. A. Mar, and B. Levine, "The Toronto Empathy Questionnaire: Scale Development and Initial Validation of a Factor-Analytic Solution to Multiple Empathy Measures," *Journal of Personality Assessment*, vol. 91, no. 1, pp. 62–71, 2009.
- [30] R. L. Reniers, R. Corcoran, R. Drake, N. M. Shryane, and B. A. Völlm, "The QCAE: A Questionnaire of Cognitive and Affective Empathy," *Journal of Personality Assessment*, vol. 93, no. 1, pp. 84–95, 2011.
- [31] D. D. Vachon and D. R. Lynam, "Fixing the Problem With Empathy: Development and Validation of the Affective and Cognitive Measure of Empathy," *Assessment*, vol. 23, no. 2, pp. 135–149, 2016.
- [32] J. Mayer, "Manual for the MSCEIT Mayer-Salovey-Caruso Emotional Intelligence Test," Toronto, ON, Canada: Multi-Health Systems, 2002.
- [33] D. D. Stevens and A. J. Levi, *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning*. Stylus Publishing, LLC, 2013.
- [34] L. D. Fink, *Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses*. John Wiley & Sons, 2013.
- [35] S. J. Lunn, C. L. Bell-Huff, and J. M. Le Doux, "Learning to Care: Faculty Perspectives on Developing Empathy and Inclusive Mindsets in Biomedical Engineering," *Biomedical Engineering Education*, pp. 1–18, 2022.
- [36] I. Mohedas, K. H. Sienko, S. R. Daly, and G. L. Cravens, "Students' perceptions of the value of stakeholder engagement during engineering design," *Journal of Engineering Education*, vol. 109, no. 4, pp. 760–779, 2020.
- [37] R. E. Wertz, N. D. Fila, K. A. Smith, and R. A. Streveler, "How Do I Understand Them? Integrating Empathy into Course Design through Personas," in *2020 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2020, pp. 1–3.
- [38] S. M. Letourneau, D. Bennett, A. Merker, S. Multani, C. J. Liu, Y. Argudo, and D. Schloss, "Narratives, Empathy, and Engineering: Creating Inclusive Engineering Activities," in *Design Make Play for Equity, Inclusion, and Agency*. Routledge, 2021, pp. 44–60.
- [39] E. A. Drost, "Validity and Reliability in Social Science Research," *Education Research and Perspectives*, vol. 38, no. 1, pp. 105–123, 2011.
- [40] R. B. Johnson and L. Christensen, *Educational research: Quantitative, qualitative, and mixed approaches*, 7th ed. Sage publications, 2019.
- [41] American Association of Colleges and Universities, "VALUE Rubrics - Civic Engagement," 2022. [Online]. Available: <https://www.aacu.org/initiatives/value-initiative/value-rubrics/value-rubrics-civic-engagement>
- [42] American Association of Colleges and Universities, "VALUE Rubrics - Intercultural Knowledge and Competence," 2022. [Online]. Available: <https://www.aacu.org/initiatives/value-initiative/value-rubrics/value-rubrics-intercultural-knowledge-and-competence>
- [43] American Association of Colleges and Universities, "VALUE Rubrics - Problem Solving," 2022. [Online]. Available: <https://www.aacu.org/initiatives/value-initiative/value-rubrics/value-rubrics-problem-solving>
- [44] American Association of Colleges and Universities, "VALUE Rubrics - Foundations and Skills for Lifelong Learning," 2022. [Online]. Available: <https://www.aacu.org/initiatives/value-initiative/value-rubrics/value-rubrics-foundations-and-skills-for-lifelong-learning>
- [45] American Association of Colleges and Universities, "VALUE Rubrics - Ethical Reasoning," 2022. [Online]. Available: <https://www.aacu.org/initiatives/value-initiative/value-rubrics/value-rubrics-ethical-reasoning>
- [46] J. L. Fleiss, B. Levin, and M. C. Paik, *Statistical Methods for Rates and Proportions*, 3rd ed. John Wiley & Sons, 2003.
- [47] D. T. Campbell and D. W. Fiske, "Convergent and discriminant validation by the multitrait-multimethod matrix," *Psychological Bulletin*, vol. 56, no. 2, p. 81, 1959.
- [48] A. Jonsson and G. Svingby, "The use of scoring rubrics: Reliability, validity and educational consequences," *Educational Research Review*, vol. 2, no. 2, pp. 130–144, 2007.