Lizzie Kahle

Using Exploratory Sequential Design to Explain the Phenomenon of Virtual Learning in Medical School and its Impacts on Professional Empathy Development: An Instrument Development Study.

Background

Clinical rotations comprise half of the medical school curriculum. Beginning in year three and through the end of year four, medical students act as both observers and interns in clinical rotation programs at nearby hospitals. Working alongside professionals in the field is the primary way students acquire the necessary skills to be a doctor. An essential component of this skillset is learning how to interact with patients empathetically: "...empathy has been described as one major element of professionalism in medicine … and the most frequently mentioned personality attribute of the humanistic physician" (Hojat et al., 2018). In May of 2020, the Association of American Medical Colleges (AAMC) recommended the suspension of all clinical rotation programs; these recommendations were held in effect until December 2022 and were enforced by almost all medical schools.

The cohort of medical students that graduated between May 2020 and December 2022 suffered enormous setbacks to their education because of the rapid shift to online learning, especially because many of the necessary components of medical school curriculum cannot be replicated in virtual learning environments. In addition to studying this phenomenon, this research study aims to also fill in the gaps in existing literature pertaining to inconsistencies found in scales designed to measure empathy. To briefly highlight the depth of this issue, a comprehensive review conducted in 2019 found that not only are there over 70 scales currently in use for measuring empathy, but also that, "nearly half of the published studies on empathy employed an empathy measure that did not align precisely with the theoretical definition the author provided" (Stosic et al., 2021). This study is therefore designed with an overarching purpose of understanding the effects of a recent phenomenon through focus groups and interviews, and then subsequently using the qualitative data to develop an instrument that more accurately assesses the subpopulations' empathic traits and abilities.

Study Design

This mixed methods study follows an exploratory sequential design. The format of exploratory sequential design format is typified by three distinct phases: qualitative, integration, and quantitative. In John Creswell's book*, A Concise Introduction to Mixed Methods Research*, he highlights the intent of an exploratory sequential design as being "to first explore a problem through qualitative data collection and analysis, develop an instrument or intervention, and follow with a third quantitative phase" (Creswell 39). When using an exploratory design in a mixed methods study for instrument development, Creswell recommends these eight steps to be

followed: review the literature, identify possible items, pretest the items with a small sample using exploratory factor analysis, conduct reliability analysis of the scales, administer the survey to a large sample, conduct confirmatory factor analysis of the results, use structural equation modeling to identify latent variables, and lastly look for evidence of construct validity (Creswell 41).

The purpose of the current study is to develop a new instrument to measure doctor empathy, therefore this study's design includes the recommended steps by Creswell as well as an additional four steps that were found to be helpful in relevant literature. In all, there are 12 total steps embedded into the three phases of an exploratory sequential design. Outlined below are the 12 steps of this study design and their corresponding phases, as well as specific details on the step's purpose and output:

Phase One: Qualitative

Step 1. Literature Review: conceptualize the construct and identify underlying theoretical and empirical principles for the construct. The intended output is to identify and select four empathy scales to use as comparison scales for reliability and validity assessments (for example, the JSE-S, EAS, IRI, SSI were the four empathy scales I focused on)

Step 2. Qualitative Data Collection through focus groups and individuals using a small, purposefully selected sample. The output is qualitative data that will be audio recorded and transcribed

Phase Two: Integration for Instrument Development

Step 3. Thematic Analysis: Stage One Qualitative Sampling Methods for Data Collection. All interviews will be audio recorded and transcribed verbatim for analysis. To facilitate data analysis, data will be entered into NVivo qualitative software. Thematic content analysis will be used to identify recurring themes that emerge in the data. First, open coding will be used to conceptualize and categorize phenomena through intensive analysis of data. An initial codebook will be developed and used to code the transcripts. Then, axial coding will be used to investigate relationships between concepts and categories that have been developed in the open coding process.

Step 4. Item generation using data from both the literature review and thematic analysis

Step 5. Conduct Exploratory factor analysis & MTMM (multi-trait multi-method analysis for determining predictive validity through assessing the convergent and discriminant validity of an instrument compared to similar and dissimilar scales)

Step 6. Reliability assessment for internal consistency (cronbach's alpha level of 0.7 or higher will indicate that the items correlate to each other and to their factor dimension)

Step 7. Pilot test the instrument and make revisions based on feedback

Phase Three: Quantitative

Step 8. Finalize the instrument into a questionnaire; administer the questionnaire to the second sample of 700 participants from 25 hospitals randomly selected through a stratification process that identified the top five leading states in both medical school graduates and practicing physicians

Step 9. Use results to conduct a confirmatory factor analysis, an MTMM analysis, and tests for construct validity

Step 10. Identify latent variables: understand underlying influences on empathy development that can not be described by the study

Step 11. Evaluate the congruence and the degree to which the quantitative results can explain the qualitative results

Step 12. (Potential future direction) Re-administer to same sample for test-retest reliability and validity, or to another large sample of equal or greater size in different states / geographical regions

The reason that an exploratory sequential design is useful in instrument development is because the first two phases (i.e. qualitative and integration) involve conceptualizing the construct, identifying theoretical and empirical foundations of the construct, developing the initial instrument through item generation, and pilot testing the instrument. Exploratory mixed methods approaches thus allow qualitative data to be translated into quantitative instruments, which in essence grant the instrument with stronger psychometric properties: more stable construct validation process "Utilizing mixed methods in various steps of the instrument development process increases instrument fidelity by assessing appropriateness with both qualitative and quantitative data…Thus, exploratory mixed methods study design [is] needed … for increasing the validity and reliability of the adapted instrument" (Shiyanbola et al., pg 799).

Through understanding qualitatively the ways virtual learning effected and altered empathy skill development for a specific cohort of medical students, this study will focus on integrating qualitative data from focus groups and interviews to create quantitative questionnaire items; the output will be a novel instrument that is developed to measure doctor empathy while taking into account the sociocultural factors of the Covid-19 pandemic, virtual learning, and a new era of doctor-patient interaction. The third and final quantitative stage is designed to test the validity of the instrument through administering a questionnaire to a large sample of the target population. The intended outcome of this study design is to engender a novel questionnaire that measures doctors' perceived empathy traits and abilities based on the reality of their situational experiences and limited educational resources for empathetic skill development.

Sampling

An exploratory sequential approach to instrument development requires that two separate and independent participant samples be established for collecting qualitative data and quantitative data. In order to ensure good integration between the qualitative and quantitative

data, the inclusion and exclusion criteria for the two samples will be the same. Two examples of the study's inclusion criteria are that the sample participant graduated from an allopathic medical school between May 2020 and December 2022, and that they currently are practicing medicine as either a medical intern or medical resident at an established hospital in the U.S. Two examples of the study's exclusion criteria are if the participant attended an osteopathic medical school and/or if they were already enrolled in a form of hybrid learning for medical education prior to May of 2020. The rationale for these two exclusion criteria are that osteopathic medical schools teach medical care, and therefore empathy, in a fundamentally different way (more holistic and family-centered approach), and these schools also do not grant the MD degree (medical doctor). The rationale for the latter is that a participant having already been enrolled in / exposed to some form of virtual learning on a regular basis prior to the Covid-19 pandemic may cause confounding bias in the results.

The total number of allopathic medical school graduates between the year 2020 and 2022 was 62,809 (www.kff.org/). Using this data on the number of graduates each year (2020: n = 20,387; 2021: n = 20,921; 2022: n = 21,051), we rounded the total value to 62,800 and used it as the estimate for our study's total target population (N = 62,800). To compare our target population to a population parameter, as of May 2023 the total number of practicing physicians in the U.S. was 1,077,115 (www.statista.com/). Therefore, the cohort that defines our target population makes up approximately ~6.0% of the total population of U.S. physicians.

The two samples used in this study should therefore share the same group membership characteristic (graduation year cohort) that distinguishes them as the target population, or as their own unique and niche "subpopulation" within the entire population of practicing physicians. However, the sample participants should be separated (i.e. no dual participants) and analyzed independently in terms of chronological time and metric criteria. As noted by Creswell on exploratory sampling procedures,"the qualitative data collection needs to be purposeful and the quantitative sample as randomly selected as possible" (Creswell 80). In accordance with Creswell's recommendations, the participant sample for qualitative data collection will be recruited using purposeful sampling methods and convenience sampling methods, while the larger sample used for quantitative data collection will be recruited using stratified random sampling and snowball sampling methods.

Qualitative Phase: Sampling Methods & Data Collection

The information collected from this group of participants will serve as the basis for our instrument. In published research that also utilizes an exploratory sequential design, most sample sizes for this phase are around 20-40. However, because this data is so influential on the proceeding steps of this study, our team decided to follow the recommendations of Russel (2002) in his article on the correct uses of factor analysis. Qualitative data will go through thematic analysis and item generation for the empathy instrument; the sample size for this group must be 100 or more in order to perform exploratory factor analysis (Russell 2002). We will recruit participants using incentives and convenience sampling. As this research study will begin in

Massachusetts, the research team will recruit a representative sample of doctors currently employed by hospitals in nearby areas. It is our belief that, because Massachusetts is ranked 10th in terms of most medical school graduates (during each year between 2020-2022 MA had ~700 allopathic medical school graduates) and 9th in number of physicians / state, our sample will be somewhat representative of the larger target population.

The participants making up the qualitative sample will be selected using purposeful sampling methods in order to obtain a small, yet representative, group of participants. An advantage to purposeful sampling is that the research team can deliberately pick a representative sample who will likely be communicative and informative for our qualitative data. For instance, because we will be conducting focus groups and individual interviews in this phase, we will use two different purposeful sampling methods: homogenous sampling and maximum variation sampling, respectively. The participants in each focus group will be determined using a homogenous sampling method so that they feel comfortable and open to sharing their experiences; the idea is that, through homogenous sampling, selecting and placing participants into specific focus groups based on similar characteristics will generate more conversation as the groups will have more in common in terms of their background and lived experiences. Maximum variation sampling method will be used to purposefully select which participants will be interviewed individually; this method will help to ensure a range of diverse perspectives and experiences.

In terms of the data collection design using this sample, the 100 participants will be placed into 10 focus groups using homogenous sampling methods. For the individual interviews, the researchers will use maximum variation sampling methods to purposefully select two-three participants from each group of 10 for a total of 20-30 individual participant interviews. The goal of conducting interviews and facilitating focus groups is to obtain perspectives from individuals and from groups to hopefully reduce social desirability bias. A disadvantage of qualitative sampling and data collection is it requires a greater time commitment from participants, which could be an issue when working with medical professionals such as doctors.

Quantitative Phase: Sampling Methods & Data Collection

This sample is meant to be larger than the initial qualitative sample and is determined using the Cochran formula. This formula determines the required sample size needed for a survey when estimating the proportion of a certain characteristic in the population. Based on the available data of medical school graduates between 2020–2022 and workforce data on current physicians practicing in the U.S. by state, we estimate that our target population cohort makes up an estimated 6% of the total population of doctors (62,809/1,077,115 = ~5.83%). In addition to the Cochran formula, factor analysis sampling is one sampling method that we use to determine the required sample size for quantitative data collection, as this group's size has to be sufficiently large enough to conduct confirmatory factor analysis of the results. Lastly, in order to reach the highest proportion of this ~6% sub-population, we will focus on administering our questionnaire to hospitals in states with the most medical school graduates and/or the highest number of

physicians. In consideration to all three of these sampling methods and statistical techniques, the required sample size for the quantitative phase is approximately 700.

In order to be representative of the target population we are attempting to develop an empathy scale for, we utilized a two-stage stratified random sampling technique. We stratified where we would administer the sample based on state data for numbers of medical school graduates and numbers of practicing doctors. Conveniently, these two stratification criteria aligned with one another; over the last three years, the same five states have consistently been in the national lead in terms of both number of medical school graduates (www.kff.org/) and number of practicing doctors. These five states are New York, Texas, California, Florida, and Pennsylvania. In order to incorporate randomization, five hospitals in each of these states will be randomly selected for a total of 25 hospitals. We will contact these hospital boards directly and offer them an incentive for helping take part in the administration of our research study, as a major component of our developed questionnaire is that it can be filled out anonymously.

Upon the cooperation of the 25 respective hospital administrators, we will send out the appropriate number of printed questionnaire copies to the hospitals, each in a concealed and discreet envelope. The questionnaire will only be administered to sample participants who share the same group membership characteristics (i.e. inclusion/exclusion criteria) as our first sample. The goal is to evaluate if the themes found in the qualitative data are representative and shared across the entire target population of doctors in this graduation cohort. The purpose of collecting quantitative sample data through written, anonymous questionnaires is to control for and/or potentially decrease the amount of social desirability bias that will inevitably occur when conducting self-report questionnaires on socially desirable traits. The only tasks we will ask of these administrators is that they handout and collect the questionnaires to the samples at the same time in order for us to control external influences. If hospital administrators/ hospital boards do not want to participate or can not comply with the guidelines and requirements of the study, we will randomly select another hospital in that specific state to take its place.

The final sampling and data collection method that is used in this phase is convergent and divergent sampling. The research team will use one the four empathy scales identified in the literature review step of phase one to be included as its own separate section of the questionnaire. Of the four scales (EAS, JSE-S, IRI, SSI) the scale that is utilized will be determined based on the results of our MTMM matrix conducted in Phase 2; the scale with the highest convergent validity scores in relation to our new instrument will be included in the final format of the questionnaire. Conversely, in order to assess the instruments predictive validity, one out of every five hospitals chosen at random will receive a questionnaire that includes an additional section with completely unrelated scale that will serve to assess the discriminant validity of our instrument.

Anticipated Challenges

After reading a number of studies on the development of an instrument for measuring empathy, the task itself seems pretty daunting. Ultimately, the best way to measure empathy would be through distinguishing empathetic traits from empathic abilities when doing thematic analysis, and subsequently developing the instrument to accurately measure for these as well as account for their differences. However, self-report traits of empathic ability are ultimately impossible to measure through Likert scales – unless the items are completely devoid of face validity measures. In the same vein, it appears that social desirability bias can be mitigated through generating items with very low face validity. However, this would be highly controversial and not necessarily "correct" in terms of psychometric properties and methodologies for instrument development. The purpose of including the sampling methods of homogenous sampling and maximum variation sampling was to hopefully reduce some sources of social desirability bias without having to resort to discounting face validity measures in totality. Likewise, the questionnaire will be anonymous and administered to the large population of participants in written format in a sealed plain envelope to create a feeling of anonymity and privacy.

In retrospect, I can confidently say that I did not choose this research topic because I had any sort of personal connection to medical education or instrument development. However, what I did find fascinating was the fact that we as researchers are only beginning to understand the ramifications of the Covid-19 pandemic. This cohort of medical students would be a key group to research for understanding just how deleterious the effects of virtual learning can be. Going forward, I would like to learn more about the statistical analyses relating to exploratory factor analysis / confirmatory factor analysis, as well as how to use the software for thematic analysis. All of these qualitative procedures and methods were completely new to me at the start of the semester; I never fully understood the level of complexity that qualitative methods and/or mixed methods can bring to research. As someone with a background in psychology and statistics, I think this is a field I would definitely like to explore further.

References

Cresswell, J. W. (2023). *A concise introduction to mixed methods research*. SAGE Publications.

Hojat, M., DeSantis, J., Shannon, S. C., Mortensen, L. H., Speicher, M. R., Bragan, L., LaNoue, M., & Calabrese, L. H. (2018). The Jefferson Scale of empathy: A nationwide study of measurement properties, underlying components, latent variable structure, and national norms in medical students. *Advances in Health Sciences Education*, *23*(5), 899–920. https://doi.org/10.1007/s10459-018-9839-9

Jamison, L. (2018, July 24). *The empathy exams*. Believer Magazine. https://www.thebeliever.net/the-empathy-exams/

Malakcioglu, C. (2022). Empathy assessment scale. *Northern Clinics of Istanbul*. https://doi.org/10.14744/nci.2022.55649

Russell, Daniel W. "In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin." *Personality and social psychology bulletin* 28.12 (2002): 1629-1646

Stosic, M. D., Fultz, A. A., Brown, J. A., & Bernieri, F. J. (2021). What is your empathy scale not measuring? the convergent, discriminant, and predictive validity of five empathy scales. *The Journal of Social Psychology*, *162*(1), 7–25. https://doi.org/10.1080/00224545.2021.1985417

Wang, Y., Yong, J., & Wu, Q. (2022). *Doctors' and patients' perceptions of impacts of doctors' communication and empathy skills on doctor-patient relationships during COVID-19*. Journal of general internal medicine. https://pubmed.ncbi.nlm.nih.gov/36253633/

Website sources for interactive population data on doctors in the U.S.:
https://www.statista.com/statistics/1270398/total-medical-school-graduates-in-the-us/
https://www.kff.org/other/state-indicator/total-medical-school-graduates/?currentTimeframe=2&sortModel=%7B%22colId%22:%22Total%20Medical%20School%20Graduates%22,%22sort%22:%22desc%22%7D