***Introduction***

The accurate inference of a target person’s thoughts and feelings as they occur within the span of an interaction (e.g., conversation, school lesson, therapy session) is called *empathic accuracy* (Hodges et al., 2014; Ickes et al., 1990; Levenson & Ruef, 1992). Empathic accuracy has been dubbed “everyday mind reading,” a process by which a *perceiver* infers a *target*’s internal states through the use of verbal, nonverbal, and contextual cues (Ickes, 2009). All empathic accuracy measures create a measure of accuracy by comparing a target’s self-reports of their thoughts and feelings to a perceiver’s inferences of the target’s thoughts and feelings. However, the three main empathic accuracy measures differ in what types of inferences a perceiver makes – i.e., what functions as the criterion for assessing accuracy. These measures are called Ickes (Ickes et al., 1990), Zaki (Zaki et al., 2008), and Likert measures (Kraus, 2017). The analyses included in this project reflect a larger line of research that seeks to understand the methodological properties of these three measures.

***Research Questions***

This project seeks to understand the degree to which each of the three main measures of empathic accuracy are impacted familiarity with the target.

***Method***

These data were collected in two phases. In *Phase I*, four 10-year-old targets (all female) were video recorded in one-on-one math lessons with their primary classroom instructor. During the lesson, the targets documented their thoughts and feelings using the three empathic accuracy measures (Zaki, Ickes, and Likert). In *Phase II*, perceivers (undergraduate students in the University of Oregon’s Psychology and Linguistics Participant Pool) viewed the four video-recorded lessons in a randomized order, inferring the targets’ thoughts and feelings utilizing the three measures of empathic accuracy concurrently (i.e., all perceivers saw each stimulus only once).

Data for the three empathic accuracy measures were mean centered. The Zaki measure of continuous affect accuracy was calculated by correlating the location of the perceiver’s slider over time with the location of the target’s slider. The Ickes measure of discrete thought accuracy was created by having four researchers code the similarity between the targets’ self-reported discrete thoughts with the perceivers’ inferences on a four-point scale (score range: 0-3, *alpha =* .87) with higher scores reflecting more accuracy. The Likert scores for discrete emotion accuracy was calculated by taking the absolute difference between target self-reports and perceiver inferences reverse scoring it such that higher values reflected more accuracy (range: 0-6). There is also a variable for time, or how much time (in seconds) a participant has had to familiarize themselves with each of the stimulus targets.

*Table 1: Uncentered descriptive statistics for empathy measures.*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mean Likert Score (SD) | Mean Zaki Score (SD) | Mean Ickes Score (SD) |
| 1 | 4.21 (0.63) | 0.36 (0.28) | 0.19 (0.44) |
| 2 | 3.73 (0.68) | 0.23 (0.46) | 0.08 (0.28) |
| 3 | 4.48 (0.66) | 0.65 (0.35) | 0.28 (0.60) |
| 4 | 4.56 (1.16) | 0.31 (0.49) | 0.21 (0.38) |

In order to improve clarity in comparing models across the three measures, all measures of empathic accuracy were POMP scored such that they ranged from 0-10, with 10 representing maximum accuracy.

A total of 204 participants were included in analyses in this study. Although a total of 275 participated, 45 participants quit prior to seeing any stimuli and 26 did not move the Zaki measure at all during the study, thus failing the attention check. In addition, of the 204 participants included in the analyses here, 99 were labeled as “semi-invariant,” meaning that they failed to move the Zaki measure during an extended period of time. This lack of movement is ambiguous: It could be that perceivers genuinely did not perceive that the target’s level of understanding varied from the midpoint of the scale throughout the clip, or, the lack of movement could reflect inattention to the target or failure to follow instructions for that clip.

To answer the main research question of whether familiarity with a target improves empathic accuracy, three models for each empathic accuracy measure were fit. The first model (M0) is a random-intercepts model with participant ID as the only predictor of the empathic accuracy score. The second model (M1) adds time as a main effect and a random slopes coefficient for time at the target level. From a theoretical standpoint, this adds to the model a random slope for each target, allowing the targets in our stimuli to vary in terms of how easy they are to familiarize to – for example, one target may be highly expressive and thus have a much higher slope than a different target, who takes more time for participants to familiarize with. The final model (M2) adds a main effect of practice to *M1*. This accounts for any order effects, such that a participant may improve on empathic accuracy generally over the course of the experiment. As an example, the Zaki empathic accuracy measure scores are plotted in *Figure 1* over time for two of the targets. As can be seen in this figure, there is considerable variability in these scores between targets.

***Results***

For each empathic accuracy measure, the three models (M0, M1, M2) were compared using AIC and BIC values as inference markers. Across the three measures, M1 reflected the best model fit, with the lowest AIC and BIC values. AIC, BIC, and R2 values for all nine models are included in *Table 2*. *Table 3* contains the model performance outputs, with bootstrapped random effects inferences. Across the three models, none had a significant main effect of time. This suggests that familiarity (or time spent) with a target does not correspond to any change in empathic accuracy across scores. However, the random effects for time nested within target contribute significantly to the model specification, with none of the 95% CI’s including 0 (see: *Table 2*). The significant limitations of these data disallow us to make inferences about the standard deviation of the random effects, but we can confidently assert that for all three empathy measures (Zaki, Likert, and Ickes) accuracy varies significantly due to targets, with targets contributing unique variance to the slopes of empathic accuracy over time. In more granular detail, the Zaki model did not have a significant main effect for time (*b* = 0.22 *95% CI [-0.08, 0.54],* but did have considerable variability in empathic accuracy scores due to participant growth (participants varied by 0.63, *95% CI* [0.53, 0.74] and due to target (targets varied by 1.68, *95% CI* [0.41, 2.96]. A similar pattern emerged for the Likert models: no significant main effect for time (*b* = -0.27, *95% CI [-0.47, 0.07],* but did have considerable variability in empathic accuracy scores due to participant growth (participants varied by 0.60 *95% CI* [0.54, 0.67] and due to target (targets varied by 1.05, *95% CI* [0.27, 2.01]. Indeed, this pattern repeated for the Ickes measures as well: no significant main effect for time (*b* = 0.09, *95% CI [-0.09, 0.28],* but did have considerable variability in empathic accuracy scores due to participant growth (participants varied by 0.30, *95% CI* [0.22, 0.36] and due to target (targets varied by 0.44 *95% CI* [0.05, 0.80].

***Discussion:***

There are substantial limitations to generalizing the inferences from the above analyses to outside samples and/or populations. First, as was described prior, there are considerable issues with invariance in this sample. Nearly half (99 of the 204 participants) failed to move the dial for at least one clip in one of the targets. This suggests considerable variability in the engagement, accuracy, and validity of the data included in these analyses. However, from a wide view, these models do contain unique and important information for future studies. First, in this paradigm, the targets contribute unique variance to empathic accuracy scores, which follows from prior research (Lewis, 2015; Lewis et al., 2012) suggesting that targets contribute key variance to empathic accuracy. Second, all three models behave similarly in that they lack a main effect for time. There is no current measure of empathic accuracy that, across targets, improves with familiarity. Familiarity is a here-to-fore unstudied contributor to empathic accuracy, and as such, this has no clear tie to the prior literature, but makes a clear case that these measures reflect more complicated construct(s) than was previously thought.

In summary, while there is no significant main effect for time in a growth model of empathic accuracy across the duration of an interaction, these results suggest that participants and targets each contribute unique variance to empathic accuracy and that the construct is more complex than can be adequately captured in a small sample size with low variability and high levels of missingness. Future research will do well to improve upon this study by further modeling the interactions between target and perceiver aspects of empathic accuracy, as well as ensure that data collected do not show similar levels of missingness as in these data.

*Figure 1: Variability of Zaki accuracy scores over time between targets 1 and 2.*

*![Diagram

Description automatically generated]()*

*Table 2:* Model Comparisons, bolded models reflect the models used in analyses.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | AIC | BIC | R2 |
| *Zaki M0* | 10798.50 | 10798.50 | 0.06 |
| *Zaki M1* | **10377.87** | **10377.87** | **0.27** |
| *Zaki M2* | 10390.66 | 10390.66 | 0.27 |
| *Likert M0* | 10480.298 | 10498.64 | 0.21 |
| *Likert M1* | **8485.529** | **8528.317** | **0.67** |
| *Likert M2* | 8492.179 | 8541.079 | 0.68 |
| *Ickes M0* | 11854.12 | 11872.45 | 0.03 |
| *Ickes M1* | **11717.47** | **11760.26** | **0.19** |
| *Ickes M2* | 11721.42 | 11770.32 | 0.19 |

*Table 3*: Model Effects

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Effect | Group | Term | Estimate | Std.Error | Statistic | Conf.Low | Conf.High |
| Zaki | fixed |  | (Intercept) | 7.61 | 0.84 | 9.00 | 5.92 | 9.15 |
| Zaki | fixed |  | time | -0.21 | 0.14 | -1.55 | -0.47 | 0.07 |
| Zaki | ran\_pars | ID | sd\_\_(Intercept) | 0.63 |  |  | 0.53 | 0.74 |
| Zaki | ran\_pars | target\_ID | sd\_\_(Intercept) | 1.68 |  |  | 0.41 | 2.96 |
| Zaki | ran\_pars | target\_ID | cor\_\_(Intercept).time | -0.99 |  |  | -1.00 | -0.87 |
| Zaki | ran\_pars | target\_ID | sd\_\_time | 0.27 |  |  | 0.06 | 0.48 |
| Zaki | ran\_pars | Residual | sd\_\_Observation | 1.84 |  |  | 1.78 | 1.89 |
| Likert | fixed |  | (Intercept) | 3.79 | 0.53 | 7.15 | 2.85 | 4.8 |
| Likert | fixed |  | time | 0.22 | 0.16 | 1.38 | -0.08 | 0.54 |
| Likert | ran\_pars | ID | sd\_\_(Intercept) | 0.6 |  |  | 0.54 | 0.67 |
| Likert | ran\_pars | target\_ID | sd\_\_(Intercept) | 1.05 |  |  | 0.27 | 2.01 |
| Likert | ran\_pars | target\_ID | cor\_\_(Intercept).time | -0.87 |  |  | -1 | 0.19 |
| Likert | ran\_pars | target\_ID | sd\_\_time | 0.31 |  |  | 0.09 | 0.56 |
| Likert | ran\_pars | Residual | sd\_\_Observation | 0.8 |  |  | 0.78 | 0.82 |
| Ickes | fixed |  | (Intercept) | 0.38 | 0.23 | 1.67 | -0.04 | 0.83 |
| Ickes | fixed |  | time | 0.09 | 0.10 | 0.92 | -0.09 | 0.28 |
| Ickes | ran\_pars | ID | sd\_\_(Intercept) | 0.30 |  |  | 0.22 | 0.36 |
| Ickes | ran\_pars | target\_ID | sd\_\_(Intercept) | 0.44 |  |  | 0.05 | 0.80 |
| Ickes | ran\_pars | target\_ID | cor\_\_(Intercept).time | -0.89 |  |  | -1.00 | 0.75 |
| Ickes | ran\_pars | target\_ID | sd\_\_time | 0.19 |  |  | 0.05 | 0.34 |
| Ickes | ran\_pars | Residual | sd\_\_Observation | 1.37 |  |  | 1.33 | 1.4 |