Accuracy Decomposition and Team Decision Making: Testing Theoretical Boundary Conditions

John R. Hollenbeck, Jason A. Colquitt, Daniel R. Ilgen, Jeffrey A. LePine, and Jennifer Hedlund
Michigan State University

The purpose of this study was to examine the implications of accuracy decomposition (D. Gigone & R. Hastie, 1997) for team decision making. Specifically, the authors tested the generalizability of the multilevel theory of team decision making (J. R. Hollenbeck et al., 1995), across various components of accuracy. The authors also tested the generalizability of this theory across different levels of staff-member specialization. Results from a study with 420 individuals in 105 teams demonstrate that the validity of the multilevel theory generalizes across specialization levels but fails to generalize across different components of decision accuracy. The authors concluded that this theory is best conceptualized as a theory of achievement accuracy, rather than mean bias or variability bias.

A great deal of research in applied psychology has focused on increasing the accuracy of decisions made by individuals and groups. Unfortunately, the accumulation of knowledge across these many studies has been limited. In a recent review article published in Psychological Bulletin, Gigone and Hastie (1997) noted that "one reason for the slow progress is that research methods and data analysis in this area are varied, difficult to compare, and often substandard" (p. 149). In particular, Gigone and Hastie cited the lack of consensus regarding the most appropriate measure of decision-making accuracy as the most serious impediment to the accumulation of knowledge in this area.

Although there are an infinite number of ways to conceptualize and operationalize decision-making accuracy, Gigone and Hastie (1997) developed a strong case for the use of mean square error (MSE) and its three components—mean bias, variability bias, and achievement correlation—as the most appropriate measure. The primary purpose of this study was to examine the affect of accepting this approach to accuracy in the area of team decision making. Specifically, we conceptually and empirically examine the degree to which one recent theory of team decision making, the multilevel theory (Hollenbeck et al., 1995), is likely to generalize across different components of MSE.

This research also examines the generalizability of the multilevel theory across different levels of staff-member specialization. The theory was originally developed as a theory of hierarchical teams with distributed expertise; however, as we will show below, there is reason to believe that the degree of specialization of staff members should not appreciably affect the major propositions of this theory. If this is the case, then this would broaden the applicability of what is currently a relatively "narrow theory" of team decision making.

Multiple Facets of Decision-Making Accuracy

Gigone and Hastie (1997) showed that almost all decision-making research has operationalized accuracy in terms of either mean absolute error (MAE; the absolute difference between the team's decision and the "true score") or the achievement correlation (i.e., $r_{xy}$, the linear association between decisions and criteria across all decisions). They argued that MSE is a superior measure of overall accuracy for three reasons: (a) it has intuitive appeal because, unlike MAE, it gives more weight to extreme errors; (b) it does not ignore the absolute differences between decisions and criteria (unlike $r_{xy}$); and (c) MSE criteria can be decomposed into three complemen-
tary parts, each of which illustrates a conceptually and empirically unique facet of accuracy. These three facets are illustrated in Table 1.

Past approaches to predicting decision-making accuracy at the individual and group levels have not decomposed accuracy in the manner described by Gigone and Hastie (1997). Thus, whereas the existing literature provides some general knowledge about variables that might be related to overall accuracy, the literature has little to say in terms of precisely specifying which variables will predict different aspects of decision accuracy. We also know very little regarding which aspects of decision accuracy are well captured or poorly captured by existing approaches. Learning which facets of accuracy are well explained by a theory is important in the theory-building process. If the variables within a theory achieve their ability to predict overall accuracy because they register on only one facet, then the best means of improving the theory is adding variables predictive of other facets. Thus, from a theory-building perspective, decomposing accuracy serves as a mechanism for managing the predictability–parsimony trade-off that exists in theory building (Bacharach, 1989; Campbell, 1991; Carnap, 1966; Weick, 1989).

This study begins to address this gap by examining the implications of accuracy decomposition on the multilevel theory of team decision making (Hollenbeck et al., 1995). We chose this theory because (a) it is representative of a growing number of conceptualizations that treat groups as information processors (see Hinsz, Tindale, & Vollrath, 1997), (b) it has received considerable support when MAE has been used as the criterion but has never been tested against MSE or its components, and (c) there are conceptual reasons to believe that the theory will not necessarily generalize across all facets of decision-making accuracy.

Overview of the Multilevel Theory of Team Decision Making

A detailed description of the multilevel theory of team decision making is provided in Hollenbeck et al. (1995, pp. 293–300). Briefly, this theory holds that team decision-making accuracy is determined by constructs that occur at one of four levels: team, dyad, individual, and decision. The theory identifies the most critical variable at each of the three lower levels and then forms aggregates of these variables at the team level in an effort to explain accuracy differences within and between teams.

At the decision level the most critical variable is decision informativeness, which is defined as the degree to which each member of the team has all the information necessary to make a decision from a given role. Decision informativeness can be aggregated to form a team level construct referred to as team informativeness, which captures how well informed a team was, on average, across all the decisions they made. At the individual level, the most critical variable according to this theory is individual validity, which is the degree to which any one staff member can generate recommendations to the leader that are predictive of the "correct" decision for the team. In teams with multiple staff members, this variable can be aggregated to form a team level construct referred to as staff validity. At the dyadic level, the most important variable according to this theory is dyadic sensitivity, which reflects the degree to which the team leader correctly weighs each staff member's recommendation to arrive at the team’s decision. In teams with multiple staff members, this variable can be aggregated to form a team level variable called hierarchical sensitivity, which captures the overall optimality of the leader's use of his or her staff.

These three concepts, decision informativeness, individual validity, and dyadic sensitivity, along with their team level analogs, constitute the core constructs of the multilevel theory. All constructs other than the six listed above are labeled noncore constructs. According to the multilevel theory, the noncore constructs are peripheral influences of team decision-making accuracy whose influence on this criterion can be attributable to their effects on the core variables (see Figure 4 in Hollenbeck et al., 1995, p. 299).

Hollenbeck et al. (1995) reported two studies that tested this theory, demonstrating that the core constructs

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Decomposition of Mean Square Error (MSE) Into Three Distinct Accuracy Facets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MSE facet</strong></td>
<td><strong>Definition</strong></td>
</tr>
<tr>
<td>Achievement correlation</td>
<td>Linear association between decisions and criteria across all decisions.</td>
</tr>
<tr>
<td>Mean bias</td>
<td>Degree to which mean level of decisions is greater than or less than that associated with the criteria (e.g., over aggressiveness)</td>
</tr>
<tr>
<td>Variability bias</td>
<td>Degree to which variability of decisions is greater or less than that associated with the criteria (e.g., extremity).</td>
</tr>
<tr>
<td>Overall</td>
<td>$MSE = (M_j - M_e)^2 + (S_j - S_e)^2 + 2(1 - r_{ey})S_eS_o$</td>
</tr>
</tbody>
</table>
explained more variance (between 25% and 50%) in accuracy than traditional variables, such as cohesiveness, familiarity, experience, attrition, job knowledge, and role redundancy. In addition, the effects of the traditional variables were mediated by the three core constructs.

Testing Theoretical Boundary Conditions

Generalizing Across Different Aspects of Accuracy

Although existing studies have been supportive of the multilevel theory, Gigone and Hastie's (1997) article was not available when the theory was originally developed. In retrospect, however, the theory seems to be one of achievement, not of mean bias or variability bias, because teams can achieve perfect levels of staff validity and hierarchical sensitivity while still exhibiting both forms of bias. For example, on a 1–7 scale, the true scores for a set of criterion decisions might include only the values 3, 4, and 5, but a team with perfect levels of staff validity and hierarchical sensitivity could accomplish this with a set of decisions that included the values 5, 6, and 7. Although the team’s decisions were perfectly correlated with the true scores and showed the exact same levels of variability (range = 2.00), there is still a substantial level of mean bias (the mean of the true scores is 4.00, whereas the mean of the team’s decisions is 6.00).

There is also nothing built into the original description of the multilevel theory to prevent variability bias on the part of teams that might achieve perfect staff validity and hierarchical sensitivity. For example, on a 1–7 scale, the true score for a set of decisions might include only the values 3, 4, and 5, but a team with perfect levels of staff validity and hierarchical sensitivity could accomplish this with a set of decisions that included only the values 1, 4, and 7. Although the team’s decisions were perfectly correlated with the true scores and showed the exact same mean (4.00), there is still a substantial level of variability bias (the range of the true scores is 2.00, whereas the range of the team’s decisions is 6.00).

Despite these potential conceptual problems with the theory, the empirical support for it has been relatively strong. The main reason for this deals with the nature of the true scores used in past empirical studies. Specifically, the nature of the decisions has been such that the true score was rectangularly distributed, and thus one could not exhibit perfect levels of achievement without a lack of mean or variability bias. That is, the scale on which all decisions have been registered has been a 7-point scale, where there was a roughly equal number of decisions that required each response (1 through 7). Thus, a team could not be off by a constant (i.e., show mean bias) and get perfect achievement (because of ceiling effects). Similarly, a rectangular distribution precludes the opportunity for variability bias, in that a team cannot achieve perfect linear consistency without using the entire range of the scale.

The research described here departs from these previous studies by using a set of decisions more conducive to generating mean bias and variability bias. In the set of decisions to be analyzed here, research participants are trained to use a 7-point scale, but all true scores are only in the 1–4 range. The reduced “true score” range and the interactive nature of the cues associated with the decision objects in this context make mean bias (specifically over-aggressiveness) highly likely. The reduced true score range also creates a possibility for variability bias (specifically extremity) because there is less variability in the true scores (i.e., range = 1 through 4) than is suggested by the scale options (range = 1 through 7). Although the lack of literature of MSE and its components precludes the development of specific a priori hypotheses, our general expectation is that the constructs specified by this theory should relate to MAE, MSE, and rxy, but not mean bias or variability bias for the reasons described above.

Generalizing Across Levels of Staff Specialization

From the outset, the multilevel theory was developed as a “narrow theory” of team decision making. That is, although there is a great deal of research on group decision making in the behavioral and social science literature, Ilgen, Major, Hollenbeck, & Sego (1993, 1995) argued that much of this research deals with people who are exposed to a common information base and have to reach consensus on a single decision where there is no objectively verifiable answer (e.g., juries). However, the teams emerging in contemporary organizations are often characterized by members who differ in terms of area of specialization and status (leader versus member), and these teams often make a number of decisions that are evaluated in terms of being right or wrong (e.g., concurrent engineering teams). The multilevel theory is narrow in the sense that it focuses on this specific type of team in this type of context.

Because specialization among staff members was a given for this theory, this factor has been held constant in all past research (i.e., all staff members have been specialized). The degree to which the theory’s relevance is limited to teams with specialized members seems questionable from a theoretical standpoint, however. That is, becoming well informed, generating valid recommendations, and appropriately weighting each staff member’s contribution should all still contribute to accurate decision making in hierarchical teams regardless of staff specialization, yet this has never been empirically established.

Thus, for the first time, the research described here tests the degree to which this theory generalizes across teams with varied degrees of specialization among staff mem-
bers. Our general expectation is that the support for the multilevel theory should not be affected by the type of team examined. If this is true, then the theory can be more broadly conceived as a theory of hierarchical teams, in general, rather than simply a theory of hierarchical teams with distributed expertise.

Method

Participants and Task

Participants were 420 undergraduates in 105 four-person teams. The TIDE\(^2\) decision-making task was used in this study (see Hollenbeck et al., 1995, pp. 301–303). TIDE\(^2\) was programmed to simulate a naval command and control scenario with a leader and three staff members. Each team member sat at a computer station that was networked to all other team members’ stations. The team’s task was to monitor the airspace surrounding the team. When any aircraft came into this airspace, each team member gathered some information about particular attributes of the aircraft (e.g., its speed, direction, angle, etc.). Each shared this information with other team members, each discussed it via written text messages, and then each arrived at a judgment regarding the appropriate response to make toward the aircraft. Judgments were forwarded to the leader in the form of staff recommendations, and then the leader registered the team’s decision on the basis of the recommendations from the staff and his or her own impression of the aircraft. Decisions were rendered on a 7-point continuum that varied in aggressiveness from ignore (1) to defend (7). The team’s decision was compared with the correct decision, which was based on translating the rules into a linear combination of the attributes and applying the equation to the attribute values of the stimulus aircraft. Teams were given feedback after each of 24 trials.

As we noted in the introduction, all previous studies testing the multilevel theory have used a rectangular set of decision objects. In the present study, although research participants were trained to use the whole 7-point continuum, the true scores for all decisions never fell outside of the 1–4 range. Thus opportunity for mean bias is enhanced because research participants, if they use the entire scale, are likely to be overaggressive. Opportunity for variability bias is enhanced, because research participants, if they use the entire scale, are likely to exhibit greater variability than is warranted in the set of decisions. This manipulation of the set of decisions did create the desired effects (i.e., the staff members showed both mean and variability bias). In terms of mean bias, the mean level of the teams’ decisions was 3.16, compared with 2.35 for the correct decision. In terms of variability bias, the standard deviation of the teams’ decisions was 1.57, compared with 1.26 for the correct decisions.

Measures and Manipulations

We operationalized the dependent variable, decision-making accuracy, in terms of both \(\text{MAE}\) and \(\text{MSE}\). \(\text{MSE}\) was also decomposed into its three components: the achievement correlation, variability bias, and mean bias. Thus, there were five different aspects of decision-making accuracy that served as the criterion in this study.

Degree of specialization was manipulated, and participants were randomly assigned to one of three conditions: a generalist condition, a specialist condition, or a general-specialist condition. In the generalist condition, staff members could measure all nine attributes of the aircraft, but they had only a general idea of the meaning on each dimension. That is, they were taught only one “critical value” on each of the nine attributes, and this critical value broke up the dimension into two gross categories (threatening and nonthreatening). Staff members in generalist teams also had to learn all three “rules of engagement.” These rules specified how values on specific dimensions combined to determine the overall nature of the threat associated with a specific aircraft.

In the specialist condition, staff members could measure only three attributes but were taught four critical values on each attribute. These four critical values broke up each dimension into five precise categories (very threatening, somewhat threatening, uncertain, somewhat nonthreatening, and very nonthreatening). This scaling reflected the actual cutoffs used to generate the true score for each aircraft and thus had higher fidelity than the system used by the generalists. Staff members in specialist teams had to know only one rule of engagement each, but this rule differed across staff members, so that as a whole, the team had each of the three rules of engagement covered.

In the general-specialist condition each staff member could measure six attributes but was taught only two critical values on each. These two critical values broke up each dimension into three categories (threatening, uncertain, nonthreatening). Each staff member in general-specialist teams had to know two rules of engagement. Although differing in terms of breadth and fidelity, the three group structures are roughly equivalent in their information storing and processing demands.

Team informity was operationalized as the total number of staff members who had all the information they needed to make a recommendation in their role. So, for example, one team member was responsible for the “motion rule” dealing with the interaction among cues dealing with speed, direction, and angle of approach. If, prior to the end of the trial, this person had all three pieces of information needed to judge the aircraft from their specialized role, they were considered fully informed. If any piece of information was missing, they were considered uninformed. Thus, this measure could take on a value from 0 (if no staff member had all the needed information) to 3 (if all did).

The correlation between each staff member’s recommendations and the correct decisions was the index of that member’s individual validity. The average of the three validities across staff members was the index of staff validity for the team.

For each team, policy-capturing techniques were used to get an index of how much weight the team leader placed on each staff member’s recommendation when arriving at the 24 team decisions (i.e., the unstandardized regression weight obtained from regressing the leader’s decision on the three recommendations). Policy-capturing techniques were also used to get the ideal weights that should have been placed on each staff member’s recommendation (i.e., the unstandardized regression weight obtained from regressing the correct decision on the
Means, Standard Deviations, and Zero-Order Intercorrelations

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Team informity</td>
<td>1.13</td>
<td>0.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Staff validity</td>
<td>0.04</td>
<td>0.18</td>
<td>0.43*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Hierarchical sensitivity</td>
<td>0.32</td>
<td>0.11</td>
<td>-0.21*</td>
<td>-0.53*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Specialist</td>
<td>0.32</td>
<td>0.47</td>
<td>0.75*</td>
<td>0.20*</td>
<td>-0.14*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. General-specialist</td>
<td>0.32</td>
<td>0.47</td>
<td>0.54*</td>
<td>-0.60*</td>
<td>0.38*</td>
<td>-0.47*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. MAE</td>
<td>1.60</td>
<td>0.30</td>
<td>-0.23*</td>
<td>-0.45*</td>
<td>0.51*</td>
<td>-0.14*</td>
<td>0.35*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. MSE</td>
<td>4.43</td>
<td>1.52</td>
<td>-0.32*</td>
<td>-0.48*</td>
<td>0.54*</td>
<td>-0.21*</td>
<td>0.42*</td>
<td>0.94*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. rxy</td>
<td>0.09</td>
<td>0.26</td>
<td>-0.30*</td>
<td>0.67*</td>
<td>-0.59*</td>
<td>-0.04</td>
<td>-0.47*</td>
<td>-0.65*</td>
<td>-0.66*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Variability bias</td>
<td>0.17</td>
<td>0.17</td>
<td>0.22*</td>
<td>-0.10</td>
<td>0.25*</td>
<td>-0.22*</td>
<td>0.13</td>
<td>0.55*</td>
<td>0.61*</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>10. Mean bias</td>
<td>0.78</td>
<td>0.63</td>
<td>0.09</td>
<td>-0.01</td>
<td>0.08</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.55*</td>
<td>0.55*</td>
<td>0.06</td>
<td>0.38*</td>
</tr>
</tbody>
</table>

Note. N = 105 teams. MAE = mean absolute error; MSE = mean square error. Specialist and general-specialist variables are dummy coded with generalist teams as the referent group.

*p < .05.

Data Analyses

All five aspects of decision accuracy were hierarchically regressed on the core constructs specified by the multilevel theory, the degree of specialization, and the interactions between these two sets of variables. The first three steps of these hierarchical regressions entered the core constructs as predictors and generally serve to test the validity of the multilevel theory. The degree to which the theory generalizes across different aspects of decision accuracy (i.e., our first research question) can be ascertained by examining the relative effect of the core constructs for each of the five different dependent variables. The next two steps of this regression enter the dummy variables capturing the type of team, where generalists are treated as the control group. The final six steps of this regression enter the interactions between core constructs and level of specialization in the team. Any significant interaction implies that the relationship between core constructs and decision accuracy depends on the degree of specialization (i.e., our second research question). All analyses were conducted at the team level.

Results and Discussion

Generalizing Across Different Aspects of Accuracy

MAE. The descriptive statistics associated with this study are shown in Table 2. Table 3 shows the regression results, which are also summarized graphically in Figure 1. The effect of the core constructs of the multilevel theory on MAE are shown in rows 1–3 and columns 1–2 of Table 3 and in the first bar of Figure 1. The results indicate that the core constructs explain a statistically significant 30% of the variance in MAE. Examination of the beta weights indicates that the nature of the effects were in the predicted direction, in the sense that MAE was smallest for teams that were well informed, had highly valid staff-member recommendations, and had a well-differentiated set of weights for aggregating multiple recommendations into a single team decision. Because this was the criterion used in Hollenbeck et al. (1995), these results replicate the previous findings associated with this theory.

MSE. The effect of the core constructs on MSE can be seen in rows 1–3 and columns 3–4 of Table 3 and in the second bar of Figure 1. The results indicate that the core constructs explain a statistically significant 36% of the variance in MSE. As was the case for MAE, the nature of these effects were in the predicted direction. Thus, these results suggest that the validity of the multilevel theory does generalize to contexts where MSE is the measure of overall decision accuracy.

In a sense, the generalizability of all of the results from MAE to MSE is not surprising. Although MSE, by its very nature, has a higher variance than MAE, the two measures generate the same rank order for teams in terms of accuracy. Indeed, despite our best efforts at pulling these two indexes apart with the specially designed set of decisions, the correlation between the two was extremely high (.94). This is not an idiosyncratic result. Even in the studies they describe in their article, Gigone and Hastie (1997) found a similar level of overlap (.84 < r < .95; see p. 156).

The true value of MSE lies not in what it provides as an overall measure but rather in the degree to which it is meaningfully decomposed into three separate aspects of decision-making accuracy. The three different facets of decision-making accuracy were largely uncorrelated (−.02 < r < .38), and our general expectation was that constructs specified by the multilevel theory would vary a great deal in their ability to predict these.
Achievement correlation. The effect of the core constructs on the achievement correlation (i.e., $r_{xy}$) can be seen in rows 1–3 and columns 5–6 of Table 3 and in the third bar of Figure 1. The core constructs explained a statistically significant 52% of the variance in $r_{xy}$. The amount of variance explained in this facet of decision accuracy is substantially larger than that explained in either overall measure (i.e., MAE and MSE). The effect sizes for all three core constructs were sizable, statistically significant, and in the predicted direction.

Variability bias. The effect of the core constructs on variability bias can be seen in rows 1–3 and columns 7–8 of Table 3 and in the fourth bar of Figure 1. As expected, the core constructs explained only a small percentage of variance in variability bias (10%), and only two of the variables (team informity and hierarchical sensitivity) were statistically significant predictors of this outcome. The nature of this effect is such that teams who were better informed and had better hierarchical sensitivity showed less variability bias.

Mean bias. The effect of the core constructs on mean bias can be seen in the rows 1–3 and columns 9–10 of Table 3 and the fifth bar of Figure 1. As expected, the core constructs failed to explain a statistically significant percentage of variance in mean bias.

Taken as a whole, these results show that the multilevel theory is best seen as a theory of achievement rather than as a theory of mean bias or variability bias. The core constructs explained a very large percentage of variance in $r_{xy}$—much more than what was found for either overall measure. The core constructs explained much less variance in variability bias, and only two of the core constructs predicted this aspect of accuracy. For the most part, staff members showed too much variability in their recommendations, but teams that were well informed and had leaders high in hierarchical sensitivity were better able to control this aspect of decision-making error relative to other teams.

Although leaders who were high in hierarchical sensitivity were able to partially remedy staffs that were too variable in their recommendations, these same leaders were not able to overcome mean bias in their staff members. In fact, none of the variables that were measured or manipulated in this study predicted this aspect of decision-making accuracy. This is unfortunate because the absolute level of mean bias was rather high in this study. That is, on average, both the staff members and the team as a whole were overaggressive by close to 1 point on a 7-point scale.
Given the large role that mean bias and variability bias have in terms of explaining variance in overall team decision-making accuracy, and given the limited ability of the multilevel theory for predicting these subcriteria, theoretical effort needs to be aimed at extending this theory by adding predictors of mean and variability bias. Because this type of bias reflects a consistent response bias, factors related to cross-situational consistency might be worth pursuing. For example, this study had no measures of personality, but certain traits (e.g., aggressiveness, impulsiveness, risk aversion) might be predictive of this type of bias.

Generalizing Across Levels of Staff Specialization

The results that pertain to the generalizability of the multilevel theory across different types of teams are shown in rows 6–11 of Table 3, where the interactive effects of staff specialization and the core variables are shown (see Figure 1 also). There were 2 (out of a possible 30) statistically significant interactions between the degree of staff member specialization and the core constructs.

The largest effect was for the interaction between hierarchical sensitivity and the specialist versus generalist distinction, which explained 5% of the variance in achievement. The nature of this interaction indicated that the effect for hierarchical sensitivity was even more pronounced in generalist teams relative to specialist teams. Similarly, the interaction between team informity and the specialist–generalist distinction, which accounted for 3% of the variance in variability bias, indicated that low levels of team informity had more pronounced negative effects on variability bias for generalists relative to specialists. Because the percentage of statistically significant findings (.07) is close to what would be expected by chance, and because the only two significant interactions were in a direction implying that the relationships predicted by the theory were actually stronger for generalists, these results do not support the notion that the validity of this theory is restricted to teams with specialized staff members.

In a separate check on this conclusion, we followed up this last test with yet another regression, where we simply tested the validity of the multilevel theory for the 37 teams of generalists. It must be emphasized that this type of subgroup analysis suffers from a lack of statistical power, because the sample size is based on 37 teams rather than 105. Despite the low power, the core constructs explained a statistically significant proportion of variance in both MAE (25%) and MSE (28%). As with the overall sample, it was also the case that, within the generalists, the core constructs registered even more highly on $r_{xy}$ (58%) than either MAE or MSE. Also, similar to the results for the whole sample, the core constructs failed to explain a statistically significant portion of the variance in either mean bias or variability bias. Thus, overall, these results indicate that although this theory was originally aimed at explaining decision-making differences among hierarchical teams with distributed expertise, it would be more appropriate to classify this as a more general theory of hierarchical teams.

References


