



Assessment of Expert Performance Compared Across Professional Domains



Rick P. Thomas*, Ashley Lawrence

Georgia Institute of Technology, United States

In this paper, we review several task characteristics to explain why experts across domains differ in their level of skill (expertise). Domains may have low levels of professional competency because of difficulty measuring relevant outcomes, impoverished performance feedback, and lack of accurate assessment tools or decision aids. Acknowledging that domains differ furthers research on expertise because it elucidates some common controversies. For example, the role of nurture (job-relevant experience) versus nature (talent or pre-existing abilities) in skilled performance, and the problem that expert-level knowledge and fast decision-making do not always translate into superior performance across domains—the process–performance paradox. Moreover, recommendations for improving domain competence must take into account the underlying differences among domains to provide recommendations appropriate for the current level of competency exhibited by the professionals in the domain.

Keywords: Professional competence, Expertise, Experts, Performance, Task characteristics

General Audience Summary

The goal of this paper is to explain why variability in performance exists between professional domains. For example, weather forecasters accurately predict next-day precipitation 82 percent of the time, whereas clinical psychologists and psychiatrists only accurately predict patient violence 39 percent of the time. We review evidence that specific professional domains may have low levels of skill because of difficulty measuring relevant outcomes, impoverished performance feedback, and lack of accurate assessment tools or decision aids. Weather forecasters have access to accurate and usable feedback for their 24-hour precipitation forecasts, whereas clinical psychologists do not have access to feedback of that quality for their predictions of violence. Understanding how differences between professional domains affect performance advances recommendations for improving competence. We also address important issues like why some professionals do not perform better as they gain experience and whether guided practice is really the only ingredient needed to develop high-level skill by evaluating the veracity of the claim that it takes 10,000 hours of practice to develop expertise. Similarly, we evaluate whether popular examples, like Shaquille O’Neal’s free-throw shooting improvement under Ed Pablashkis and Judit Polgár’s skill development in chess under her father’s tutelage, are truly mentorship success stories.

Author Note

We would like to thank Andrew Butler and two anonymous reviewers. This article was supported by the Office of Naval Research (N000-16-2863).

* Correspondence concerning this article should be addressed to Rick P. Thomas, School of Psychology, Georgia Institute of Technology,

654 Cherry Street, Atlanta, GA 30332-0170, United States. Contact: rick.thomas@psych.gatech.edu

This paper reviews the literature that indicates experts in different domains vary in their degree of professional competence. We will consistently use the term professional competence to refer to performance across experts at the domain level, expertise as individual level performance, and reserve the term expert to refer to the professionals in a domain whether they exhibit expertise or not. We review a few critical task characteristics that potentially explain why experts in different domains exhibit such varied levels of expertise. In other words, differences in task characteristics are an important reason why there is so much variance in professional competence. Based on our review, we make several recommendations for advancing the cognitive science of the study of expertise and suggest strategies to improve professional competence.

Domain Differences in Professional Competence

Researchers have relied upon several techniques to identify and assess the professional competence of different domains. [Table 1](#) operationalizes these techniques and their major weaknesses ([Shanteau, Weiss, Thomas, & Pounds, 2002](#)). Note that these procedures are unnecessary in domains that have good validity criteria—correct answers are assumed to be known and uncontroversial.

In many tasks within professions that do not have clear and valid criteria, both consensus and consistency are important metrics of professional competence. [Einhorn \(1974\)](#) argued that domain experts should agree with each other (consensus) and themselves when making repeated judgments about the same cases (consistency). If the biopsy is judged by one expert to be cancerous and by another to be benign, we should probably worry about the level of professional competence in such a domain. Consistency criteria are perhaps even more important. If an expert judges the biopsy slide as indicative of cancer this week but had determined the identical slide was benign last week, we should consider this *prima facie* evidence against the diagnostician's expertise. Self-consistency is a necessary condition for expertise in many professional domains and if the experts exhibit low consistency (intra-rater reliability), poor consensus (inter-rater reliability) follows because consistency is a necessary condition for consensus. Thus, if experts exhibit inconsistency in tasks where it is reasonable to assume a single correct answer, at least in theory, then it seems prudent to question the professional competency of the domain.

[Shanteau \(1992\)](#) argued that experts across different domains exhibit widely varied levels of professional competence. In a recent update to his 1992 seminal article, [Shanteau \(2015\)](#) reported estimates of consistency and consensus across several different domains (recreated in [Table 2](#)). The consistency and consensus of experts across different domains varies dramatically. Stockbrokers and clinical psychologists exhibit relatively meager intra-rater reliability. Weather forecasters and auditors, on the other hand, exhibit relatively high levels of intra-rater reliability. Thus, there seems to be little doubt that different domains of expertise vary considerably in their professional competence. But the primary question of interest is why?

The Role of Task Characteristics in Professional Competency

Importantly, [Shanteau \(1992, 2015\)](#) speculates that task characteristics drive the differences in professional competence that manifest. In the following subsections, we will discuss some of the most important domain differences that appear to influence professional competence.

Challenges to Validating Predictions

Professionals in many domains are asked to predict extremely noisy, low base -rate, and often ill-defined criteria. In general, experts in domains that regularly forecast human behavior typically perform poorly ([Table 2](#)), likely because human behavior is fantastically difficult to predict. Although actuarial models predicting recidivism account for substantially more variance than human professionals, the best statistical prediction rules typically exhibit fair performance for predicting future arrests and are poorly diagnostic for predicting violent crime ([Zeng, Ustun, & Rudin, 2017](#)). This result should be sobering because it suggests that a great many outcomes that society cares deeply about (a felon's proclivity to commit a violent crime after being paroled) are extremely difficult to predict.

Human behavioral outcomes are often difficult to assess and fraught with measurement error. For example, one goal of a clinician is to determine whether or not patients are likely to hurt themselves or others in the future. Measuring whether or not harm has occurred after the initial prediction is difficult because it relies on future contact with the patients and the accurate self-reports of the patients. On the other hand, in good-performing domains it is typically the case that measures of relevant outcomes can be made with little error and effort on the part of the experts themselves (e.g., determining whether it rained or not last night). Thus, it is not surprising that a large portion of the poor-performing professions on the right side of [Table 2](#) tend to evaluate human behavior.

Substituting Inadequate Gold Standards

An issue related to the difficulty of measuring criteria is the concept of a gold standard—an agreed -upon measure used in place of the valid criterion (e.g., using ejection fraction, which measures damage to the heart, as a surrogate for whether someone had a heart attack). Some gold standards do not have much error and are relatively well -defined. The problem, however, is that gold standards are often substitutes for the criterion that the professionals are actually trying to predict and may not accurately reflect the criterion. Returning to recidivism, what parole boards want to predict is whether a criminal will commit a violent crime if let out of jail. However, the crime that counts toward recidivism may not necessarily be indicative of the violent behavior the parole board was trying to predict. Moreover, whether the parolee gets caught is partially reflective of policing practices ([O'Neil, 2016](#)). Thus, the gold standards substituted for many criteria of interest may not adequately measure the construct being evaluated and predicted in many professions.

Download English Version:

<https://daneshyari.com/en/article/7241585>

Download Persian Version:

<https://daneshyari.com/article/7241585>

[Daneshyari.com](https://daneshyari.com)