Proficiency Scaling of Warning Forecasters

Daphne LaDue, Center for Analysis and Prediction of Storms, University of Oklahoma
Pamela L. Heinselman, NOAA National Severe Storms Laboratory
Robert R. Hoffman, Florida Institute for Human and Machine Cognition

Motivation

- New radar technology has the potential to better capture storm processes, allowing forecasters to connect conceptual models of storm processes to depictions in data.
  - Do these depictions help forecasters improve performance? Can journeymen build proficiency faster? Does the new technology hamper the expert?
- Data from multi-function phased array radar helped three teams outperform three other teams during PARISE 2010 (Heinselman et al. 2011).
  - Some teams clearly had more experience with tropical supercell and tornado evolution.

PARISE 2012

- 12 forecasters each individually worked 4 cases
- Research protocols captured forecaster thinking and reasoning
- Dataset contains detailed information on how and when key judgments were made, and use of conceptual models
- This portion of our work focuses on identifying forecaster proficiency for use in other parts of the analysis

Twelve Participants

Central and Southern Regions of the NWS

Time in Service	Highest Degree Held	Role in the Office	Sex
• 4 @ 1–3 yrs • 2 @ 5 yrs • 6 @ 9–20 yrs	• 9 B.S. • 3 M.S.	• 2 Interns • 3 Journeymen • 2 General Forecasters • 1 Warning Coordination Meteorologist • 4 Senior Forecasters	• 9 males • 3 females

Describing Proficiency

Below are complementary ways to describe a forecaster's proficiency: by achievement level, or by the underlying cognitive strategy behind the achievement.

***Proficiency categories**, adapted from Hoffman, 1998. These categories were extended to work of the professions from the terms used by craft guilds. Three are relevant to the NWS forecaster, who has completed formal schooling requirements.*

Apprentice: Someone immersed in the domain, and learning beyond the introductory level. Lengths of apprenticeships vary from 1–12 years.
Journeyman: Someone who can work unsupervised, although under guidance. Experienced and reliable. Competent. This could be the highest level someone attains.
Expert: Someone with uncommonly accurate and reliable judgments, highly regarded. Extensive experience in the subdomain has led to high skills and specialized knowledge.
**These are commonly subdivided. For example, junior journeyman – journeyman – senior journeyman.*

***Cognitive style designations** of Pliske, Crandall, and Klein, 2000. These categories were built in their studies of weather forecasters. The levels that appear to apply best to NWS forecasters are included here and are being used in this study.*

Scientist: Experts. Adapt/modify strategies easily to the problem of the day. Constantly self-assess their strategies and reasoning. They engage in "what if?" scenarios to anticipate problems. "Reflective practitioners."
Proceduralist: Journeymen. Fixed set of strategies applied to challenge of the day. Do not think about or question their strategies or assumptions. Follow procedures closely. Some may stay here, but some will advance.
Mechanic: Always follow standard procedures. Superficial knowledge. Do not care much or wonder about underlying concepts. Want only to accomplish the job; not motivated to improve.

Applying Methods to Determine Proficiency

Proficiency Scaling is a process designed to identify the experts, journeymen, and apprentices in an organization to inform development of training, operational software, etc.

Proficiency scaling should be based upon at least two of three methods. We are using all three to some extent. They are presented in the order in which we are seeking them:

Sociometric Determination

1 **Ratings based upon sociometrics***
We requested two contrasting groups during participant recruitment: 1) journeymen within one year of taking DLOC, and 2) some of the best warning forecasters in the Southern and Central region.

In its simplest form, a sociogram could focus on a single relation: whom do you trust? We began here, during recruitment, to find our expert group. Our journeyman group was more simply requested as "forecasters within one year of having taken the Distance Learning Operations Course." DLOC is the NWS's initial training course on use of WSR-88D radar.

**There are more extensive forms of this type of scaling.*

Quantitative Performance Measure

2 **Quantitative performance measures**
Verification scores for PARISE are shown below. We will be comparing these with their recent operational verification scores to establish how performance during PARISE varied from normal performance.

11 May 2010 EF0 tornado:

- 92% of PARISE 2012 lead times exceeded the more stringent 18-min national average lead time.

First of three tornadoes during the 22 May 2011 case:

- 58% of PARISE 2012 lead times exceeded 18 min.

These figures are more fully explained in Heinselman et al. 2011 (26th Conference on SLS, October 2011).

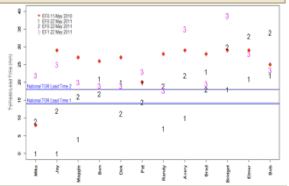


Figure 1. Distribution of tornado lead times (min) computed for 11 May 2010 and 22 May 2011 events: EF0-rated tornado on 11 May 2010 (red dot) and 3 tornadoes on 22 May 2011; EF-ratings listed in chronological order. Lead Time 1 (2) is the average national lead time (horizontal blue line at 14 (18) min) resulting from the inclusion (exclusion) of tornado events with negative (set to 0 in TLT computation) lead times. *Source: Heinselman et al., 2011*

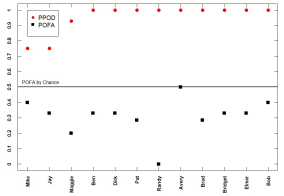


Figure 2. Distribution of polygon probability of detection (PPOD) and probability of false alarm (POFA) computed for 11 May 2010 and 22 May 2011 events. The horizontal line at 0.5 indicates the POFA attainable by chance.

Extent, Breadth, and Depth of Experience

3 **Estimates of extent, breadth, and depth of experience**
Our Career Interview tool, based on previous work by Hoffman, was the primary basis. Recent Case Walk-Through and Leveraging Knowledge tools are also being used.

The 10-year rule-of-thumb of the experience needed to become an expert is a rough proxy; Pliske et al. found that many military forecasters never moved past being a "proceduralist." Alternately, the storm chasing or weather enthusiast may have acquired some mastery in, for example, radar data interpretation outside their education.

The following tools were designed to more deeply probe and identify participants' expertise level, and factors which led to their current level.

Career Interview *Structured Interview*
Documented early interest, education in/beyond formal schooling, career steps, and duties of their current position.

Recent Case Walk-Through *Narration of Desktop Video, Followed by Structured Interview*
Guided the forecaster through recall of cognitive actions and identification/explanation of judgments made during the warning case just worked. Interview questions then focused on information used, aspects of the decisions made, whether and what conceptual models were used, and how work strategies deviated from their normal or taught strategies.

Leveraging Knowledge *Structured Interview*
Sought instances and stories that helped define what individual forecasters saw as their key work strategies for tornado warnings, what they find difficult versus easy, what defines typical versus innovative ways to accomplish the job, and what they see as deviations in weather from what was expected. They were also asked to describe instances where they had insight into how an event was evolving that others did not catch.

Pliske et al.'s Cognitive Activities

Numbers indicate frequency of how Pliske et al. coded categories in their data (narratives of weather forecasts), which they subsequently organized by forecaster skill level. We will apply this technique to our warning narrative data.

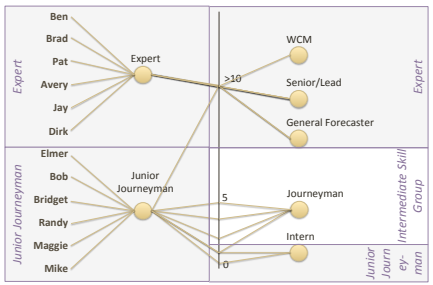
Cognitive Activity	High Skill	Medium Skill	Low Skill
Noticing Patterns Comments on co-occurrences, deviations, and/or patterns in weather information.	3.8	3.5	1.5
Seeking Information Actions taken to collect information to produce a forecast.	3.8	2.7	3.0
Meaning Making Process of organizing and explaining weather information to make sense of it.	5.0	2.8	1.5
Visual Mental Representation Use of visual mental representations when discussing forecasting processes.	1.2	1.2	0.2
Metacognitive Processes Self-reflection	1.4	1.2	0.8

Factors that appear to have provided the best differentiation between skill level for general weather forecasting in Pliske's work:
**Seeking Information*
**Meaning Making*

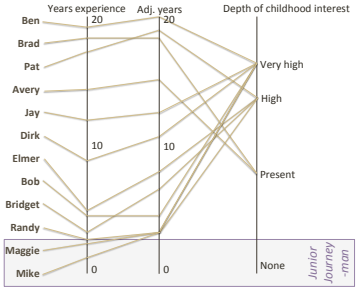
Factors that initially indicate the best discrimination ability between skill levels (subjective) in this work:
**Metacognitive Processes*

Case Walk-Through narratives suggest that the warning decision process is a far more specific and intense analysis of data. Initial coding indicates far higher incidents of these cognitive actions. In Pliske et al.'s work, Meaning Making best discriminated between skill levels. However, even young NWS forecasters are verbalizing the meaning of the signatures, as they are taught in DLOC. A sharper discrimination between education and developed expertise will be necessary for this study due to the more intensive and quality training NWS forecasters receive in radar interpretation and warning decisions.

Iterations in Identifying Proficiency (To Date)



1 Recruitment
Refine: Years since DLOC
Refine: Role in Office



3 Years Experience
3 Childhood Interest

Immediate problems: one of our first junior journeymen participants was 5 years past DLOC and had extensive storm chasing experience.

Only three participants were within a year of having taken DLOC.

See #3 above. "Adjusted years" includes all employed forecasting-related experience (SCEP, other forecasting job).

Childhood interest may effectively add to years experience, depending on how they pursued their interest.

Conclusion: These simple measures suggest we did not achieve two strongly contrasting groups during recruitment. Some of these considerations, however, are less specific to current skills and more informative toward future skill level potential (e.g. deep childhood interest may indicate long-lasting motivation to develop deeper understanding). Analysis of cognitive strategies used during the cases will better reveal adaptability in thinking and depth of understanding.

Next Steps

The more intensive portion of this work will be finished this spring, when coding of walk-throughs and leveraging knowledge documents will be completed.

We will also establish benchmark interpretation and warning performance to allow interpretation of forecaster assessments of each storm.

Analysis of procedures taught in forecaster training (DLOC and AWOC) will enable identification of deviations from trained strategies. Forecaster comments on office practice will enable identification of office or regional strategies.

Beginning stages of the more intensive analysis indicate that some forecasters adapted their strategies according to their assessment of the key problems in each case while others did not. The more intensive analysis should better discriminate current skill levels.

While applying Pliske et al.'s scheme to warning forecasters we will also evaluate and refine its applicability to warning (as opposed to forecast) decisions.

These steps will inform interpretation of quantitative performance measures and establish whether and how PAR technology aids both experienced and inexperienced forecasters.

We thank the 12 NWS forecasters for their participation in this study, the Southern and Central Region SSD Chiefs, MICs, and SOOs for aiding recruitment, James Correia for assisting in data collection, and Les Lemon and Steve Martinaitis for their participation in our experiment test run. We also thank the following colleagues for their contributions to this research. Experimental Warning Program leads Greg Stumpf, Travis Smith, and David Andra; A/V specialist James Murnan; GIS expert Ami Arthur; NWS verification expert Brent MacAloney II; WDSS-II expert Kiel Ortega; simulation expert Dale Morris; software experts Eddie Forren and Hoyt Burcham. This work was prepared by the authors with funding provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA11OAR4320072, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of NOAA or the U.S. Department of Commerce.