Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Purchasing & Supply Management 10 (2004) 211–222
JOURNAL OF PURCHASING AND SUPPLY MANAGEMENT
www.elsevier.com/locate/pursup

The myth of purchasing professionals' expertise. More evidence on whether computers can make better procurement decisions

Frits Tazelaar[a], Chris Snijders[b,*]

[a]Department of Sociology/ICS, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands
[b]Department of Technology Management/ECIS, Subdept. Of Technology and Policy, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Received 28 June 2004; received in revised form 1 September 2004; accepted 20 November 2004

Abstract

In a previous experiment, we have shown that risk assessments of purchasing experts are certainly *not* better than that of subjects untrained in purchasing, and worse than the decisions made by formal models (J. Purchas. Supply Manage. 9 (2003) 191–198). Since both these results are rather counterintuitive, we conducted a series of experiments geared at replication and extension of these findings. These new experiments show that our previous results are robust, and reveal an additional finding that is both worrying and puzzling. It actually seems to be the case that for the purchasing decision tasks in our experiments, experts perform worse with growing experience. It therefore seems that, at least for the kinds of purchasing decisions under study, it does not make much sense to use expert judgments at all. However, we show that there is a way in which expert judgments can be used in combination with formal models to improve the predictive accuracy of purchasing predictions. In our case, superior predictions are made when we combine the prediction of a formal model with the prediction of the 'average expert', thereby combining the robust linear trends as encapsulated in the formal model with the more intuitive configural rules used by experts. We provide several explanations for this phenomenon.

1. Introduction

Most people would agree that at least one of the tasks of a purchase manager is to decide which of a set of purchasing transactions needs purchase management more. For instance, for some transactions it makes sense to ask for many tenders, invest much in the screening of suppliers, involve much time in negotiating, and put a serious effort in writing a detailed contract. For other transactions such investments are not effective or not efficient (Batenburg et al., 2000).

Although it is typically part of a purchase manager's job, there are compelling arguments on the basis of the literature on *clinical* versus *statistical prediction* that suggest that purchase managers—like all other humans—are typically not good at making precisely these kinds of judgments. In a review study, Grove et al. (2000) have shown that for single, quantitative decision tasks computer models almost always perform at least as good as or better than human experts. Most of the studies they reviewed were based on tests in medical, forensic and clinical-personality studies (102 out of 136). There are only a handful of studies comparing human experts with models that deals with 'more economic topics'. Grove et al. (2000, Table 1) mention studies on business failure, job performance, job turnover, business startup success, job success and work productivity. In a previous publication (Snijders et al., 2003), we set out to test this assertion when it comes to judgment and decision-making in purchasing, and reported on an alarming and somewhat counterintuitive result. It indeed turned out that, for the cases under study,

*Corresponding author. Tel.: +31 40 247 2640; fax: +31 40 246 4646.
E-mail address: c.c.p.snijders@tm.tue.nl (C. Snijders).

1478-4092/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pursup.2004.11.004

Table 1
Spearman correlations between actual and predicted scores, averaged per group, as in Snijders et al. (2003)

Formula	0.37
Students	0.26
Purchasing professionals	0.24

purchasing professionals do *not* make better purchase decisions than undergraduates, and both are actually outperformed by a computer model (using a simple formula).

Though in principle this result is in line with research in other areas, and in that sense perhaps need not be treated as a surprising finding, we want to analyze some of the evident follow-up questions related to this result. First, we consider how robust this finding is. Perhaps our finding was simply for a purchasing professional rather unlucky statistical fluke, so we tried to replicate our findings (and succeeded). Second, we extended our previous research and considered possible reasons for our findings by looking at the way in which experts, in general, tend to behave. Finally, we focus on whether and how the decision-making of purchasing professionals can be improved, based on the literature on experts and expertise.

2. Computer beats purchasing professionals: a robust finding

We first repeat in brief the essential elements of the experimental setup as used in Snijders et al. (2003). In that experiment, both purchase professionals and undergraduates were each given 8 case descriptions regarding a procurement transaction (the procurement of IT-products; see Fig. 1 for an example case), and were asked, among other things, to predict the likelihood of this transaction being a problematic one (see Snijders et al., 2003 for details). All our claims about the judgment and decision-making capabilities of purchasing professionals are, therefore, based on purchasing professionals being able to identify which transactions are the ones most likely to be problematic. This is, obviously, not the only kind of decision a purchasing manager has to make, but we think it is an important one. Moreover, the purchasing professionals themselves felt they would do fine in this task, even when asked after completion (but before displaying the results).

In fact, the vignettes were chosen from a larger database of real purchasing transactions, so that we actually knew the correct answers to what the purchasing professional and the students were predicting and could compare their answers with the real ones (for a more detailed description of this database, see the Appendix A, and Batenburg, 1996 or Buskens and Batenburg, 2000). In short, our professionals and students were to predict how many problems would occur, and we knew the right answer. All undergraduates involved were freshmen in information sciences and participated as part of a course requirement. Purchasing professionals participated in reaction to an invitation from a student. Each pair of students had to find one purchasing professional who was willing to participate. Preferably, the professional should have experience in both purchasing and IT, since the transactions on the vignettes were all about IT-products. Ultimately, 30 purchasing professionals and 60 students participated.

We made sure that for (almost) all sets of eight vignettes, there were three individuals who made predictions: a purchasing professional and two students. The 240 (30×8) vignettes given to the purchasing professional were all different. This guaranteed a large spread in the kinds of vignettes under consideration and it enabled a clean comparison of purchasing professionals versus students. Participants were also asked to answer several other questions to which we will return later.

Beforehand, we calculated a formula that generates predictions on the likelihood of a particular transaction being problematic (prediction 1) and which kinds of problems were to be expected (prediction 2, for four kinds of problems). No fancy modeling was used: for both issues the formula was linear in the predictors. For instance, the formula to predict the likelihood of problems was a linear combination of the price of the product (if high, then more problems), the importance of the product for the profit of the buyer (if high, then more problems), the buyer's ability to judge the price/quality ratio of the product (if high, then more problems), and the degree of detail in the written contract (if high, then more problems). Note that this , roughly shows that the model simply expects more problems for larger and more complicated transactions, irrespective of the amount of effort involved in careful planning and the kind of partner.[1] We did not include the kind and number of products as a predictor for the model; this even gives a small information deficit to the model as compared to the respondent.

Hence, we used five formulas: one to predict how problematic a transaction was going to be, and four separate formulas, one for each specific problem that could occur. Since we calculated the formulas on the

[1] More careful and elaborate measurement and data analysis using the dataset as mentioned in Buskens and Batenburg (2000) show that other aspects than just these four also correlate with the amount and likelihood of problems in a purchasing transaction (cf. Rooks, 2002). That the model to predict problems as used in this paper is relatively straightforward is in part a consequence of the fact that one only has 14 separate items available for prediction.