

Herbert Test Helping Hire Software Developers: A Complex Algorithmic Problem-Solving Tool

Soraya Cardenas, Ph.D.
Sociology Department
Cascadia College
Bothell, WA (USA) scardenas@cascadia.edu

Abstract—Companies that hire software developers face challenges hiring employees. As the cost of hiring developers can cost up to hundreds of thousands of dollars, depending on expertise of the desired candidate, companies must rely on a multi-method approach to secure competent developers. This paper evaluates the “Herbert” test, an algorithmic problem solving program that is currently being used to help evaluate potential software developers at Fast Track, a web development company. There appears to be a positive correlation between performance on the test and subsequently on the job. This longitudinal study reaffirms the need for a multi-method approach to hiring with a special emphasis on problem-solving. Furthermore, as technical systems continue to emerge, increasing the demand for technical jobs, testing tools like “Herbert” are becoming more relevant.

Keywords—Algorithms; evaluations; problem-solving and hiring software developers

I. INTRODUCTION

In the United States, the field of Software Engineers is growing at about 17 percent [1]. The cost of hiring a software engineer is not cheap, averaging \$23K-\$32K for the recruiting process alone [2]. There is an investment in the hiring process of vetting the candidates and once the company chooses their preferred employee, they must contend with companies trying to lure their talent away with higher salaries. It’s not unheard of for software developer’s right out of college to receive salaries up to \$90,000; only to receive an additional 20 to 25 percent bump for accepting another offer [3]. In addition to exorbitant salaries, companies must provide health care and other amenities, which add another \$30,000 to the pot [4]. Companies further contend with high turnover rate, resulting from disgruntled employees to ill-equipped talent [5]. Often companies will be forced to hire outside the United States to help meet their demands, but this could be complicated as the quality of education and skill set could vastly vary. According to a report from India, only about 7 percent of software developers graduating from India universities are considered employable because India’s educational systems are not equal across the country; thus, those who can afford a top education receive proper training [6]. However, the real cost in hiring developers comes at a risk of hiring unqualified developers. Bad software developers can decrease team’s productivity, increase software maintenance and

troubleshooting costs, and even harm firm’s reputation – possibly causing hundreds of thousands of dollars or more of damage [7]. This makes hiring qualified software developers critical in controlling costs and in business success. Thus, when companies hire software developers, they need to rely on multiple approaches to vet their applicants, from reviewing resumes, phone interviews and onsite visits to secure viable and competent software developers. During the onsite visit, candidates may take assessments that analyze their coding abilities and “algorithmic development, data structures and analytical thinking” [8]. This is coupled with intensive interview questions, asking candidates to solve real problems that they may encounter on the job. Applicants also must engage with a stream of employees, testing their ability to interact with others [8]. Companies appear to be addressing three significant areas: knowledge of discipline, problem-solving and personality attributes. Though these three characteristics are important considerations when hiring, studies have demonstrated that attributes associated with cognitive functions appear to be one of the highest predictors of employee success, specifically critical thinking and problem-solving [9], [10].

In this longitudinal study, I evaluate the Herbert test and its “h” language, developed by Brian Conte, CEO of Fast Track, and a company that focuses on web development. According to Conte, “h” language is a language that uses algorithmic principles to solve puzzles (Fig. 1), which range in complexity [11], [12]. According to Conte, Herbert was used to assess the problem-solving skills of over 40,000 students in Microsoft’s Imagine Cup’s world-wide algorithm competition for several years and for over 17 years at his Fast Track company to test entry level software developers. This study will analyze 12 years of data collected, evaluating whether there is a correlation between high Herbert scores and software developer success.

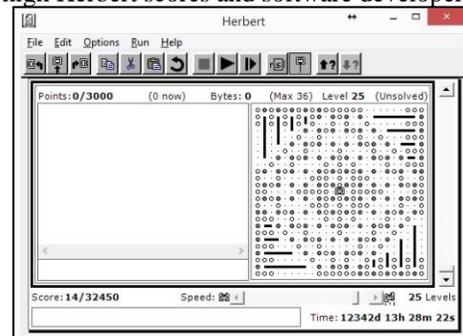


Fig. 1. Example 1 Herbert test puzzle.

II. RELATED WORK

In an article by Hewner and Guzdial (2010) [13], they administered an online survey to a game developing company with over 100 employees. Out of them 32 employees responded. The survey attempted to extract what was considered the most important attributes when hiring college graduates for game development positions. What they found were three significant areas: 1) personality, checking your ego at the door; 2) knowledge of coding and data structures; and 3) solving algorithmically challenging problems [13]. These results again resonated in a later study by Li, Ko and Zhu (2015) [14]. They conducted 59 semi-structured interviews with seasoned software engineers, some with over 25 years of experience, across 13 divisions at Microsoft to determine what key characteristics of successful software engineers were. They honed in on 53 attributes and divided them into four categories: personal characteristics and decision making, which they labeled as internal attributes, and teammates and software product, which they labeled as external attributes. Under decision making, they focused on four areas: 1) knowledgeable about people and the organization: this related to developers' ability to understand their co-workers' responsibilities, knowledge, and tendencies; 2) sees the forest and the trees: this related to developers seeing the big picture and extrapolating it down to multiple levels; 3) update their mental models: this related to developers' ability to adapt especially with changing technology; and 4) handling complexity: this related to developers understanding and ability to address complexities related to problems [14].

Decision making process could be categorized under problem-solving characteristics, especially with handling complexities. Problem-solving also resonated with the previous study on solving algorithmically challenging problems. Thus, a recurrent theme in finding successful developers is problemsolving, which can be divided into two modes: convergent and divergent, which is basically analytical and creative problemsolving. Convergent problem-solving refers to solving problems that are rational and well-defined. There is usually a correct answer, such as $1 + 1 = 2$. Meanwhile, divergent thinking leads to a large quantity of ideas that are not necessarily correlated, for example, $1 + 1 = 1$ when applying alternative algebraic principles [15]. When further analyzing the concept, problem-solving, researchers found that this was another method to measure reasoning skills in IQ testing [16], which research has shown that there is a correlation between cognitive function and success, meaning better problem

solving skills, the greater likelihood of success in one's career [9], [10], [17]. Thus, when evaluating the data obtained through Fast Track, the assumption was made that those who had higher scores in "h" language would fare better in regards to "hire-ability" and performance.

III. METHOD

This study used 12 years of data collected by Fast Track on their use of the Herbert test in the interview process, an

algorithmic problem solving tool. The Herbert test requires the user to learn and use a simple computer language called "h".

"h" uses traditional high-level language concepts: statements, procedures, parameters, arguments, and recursion (Fig. 2); "however, "h" is syntactically more simple, and contains some concepts (procedural arguments) that are not found in traditional languages" [11]. The objective of the test is for the participant to solve as many of the 10 puzzles as possible within the allotted time. The puzzles begin with a robot, which is called Herbert (Fig. 1). He can only move left, right and straight. Participants must program the robot to hit all the white buttons while avoiding the gray ones. In order to "solve" a level, the program must work and also be optimized to fit within a certain number of "bytes". Points are awarded based on both the correctness as well as the optimization of the solution. For example, if Herbert must move forward 5 spaces and the puzzle has only allocated 5 bytes to complete the solution, then using the statements ssss would meet the criteria to hit the desired result. What if the participant types a program that utilizes recursion?

```
a:sa a
```

The puzzle has been solved in 4 bytes; thus, the participant receives additional points for improving upon the intended solution. The "h" language test encourages participants to find better solutions, reinforcing optimization principles. Participants receive additional points for finding better solutions. Points also are based on the increasing complexity of the puzzles, so the harder the puzzle the more points the participant receives. Each participant has 30 minutes to complete the test, but prior to the evaluation, participants spend 15 minutes on a tutorial (Fig. 2).

A standard Herbert test was given to 3,880 software developer candidates since 2003. This test was administered as part of the candidate selection process, which also included a general IQ test and personal interviews.

Of the 3,880 candidates who took the test since 2003, 127 were hired, with their job performance subsequently tracked. The data set used for this study consisted of scores from the most recent 556 candidates, whether selected or not, and the scores and subsequent job performance of all of the 127 hired engineers over the years. This study looked for correlative relationships between Herbert scores and employee "value" at three points in their job tenure:

(<0.0001%). This result is almost certainly significant.

Another way to state the results is that the Herbert score of selected candidates were, on average, 37% more than those not selected (Fig. 3).

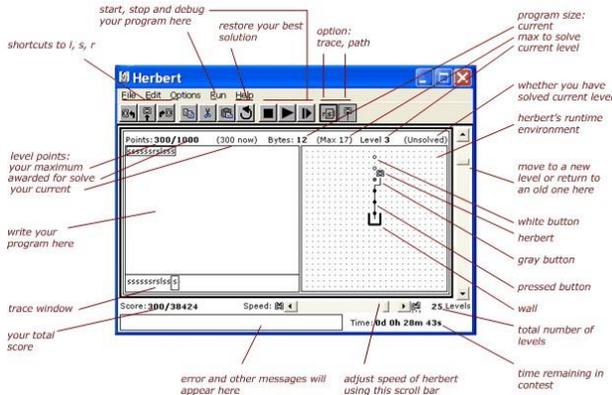


Fig. 2. Herbert test tutorial sample.

- 1) At time of candidate selection.
- 2) After their second performance review (approximately 1 year after being hired).
- 3) Over the lifetime of their career with Fast Track.

For (1), the study looked for a correlation between their Herbert score and whether or not they were given a job offer.

For (2), the study looked for a correlation between the candidate's initial Herbert score and the raise they were given on their second review. The second review was used as raises are rarely given during an employee's first review at Fast Track as the employee is usually still considered in a training or trial period, so it was felt that the raise awarded in the second review was a better indicator of initial employee value.

For (3), the study looked for a correlation between the candidate's initial Herbert score and their overall "value rating" as given by the Development Manager at the end of their tenure, or for ongoing employees, throughout their tenure to date. Note that the Development Manager was the same person throughout the period of the study, so these scores were inherently normalized.

This study also removed any identifying markers from the data set to protect the developer's identity.

IV. RESULTS

Positive correlations were found between a candidate's initial Herbert score and all three indicators of job value, with the significance level of these decreasing over a candidate's tenure with Fast Track.

Table 1 summarizes the correlations found between the candidate's score and the investigated outcomes, as well as the P-value of the results at a 95% confidence level.

A. Candidate Selection

Here the study correlated the candidate's Herbert score against whether or not they were selected to be hired.

The study found a 16.5% correlation between Herbert score and whether or not a candidate was selected to be hired, with a sample size of 1105. This gives a very low P-value

TABLE I. SUMMARY OF CORRELATIONS

Results	Correlation	Sample Size	P-Value (0.05)
Candidate Offer	16.5%	556	0.000092
First Year Raise	19.3%	93	0.06
Manager Rating	12.7%	127	0.15

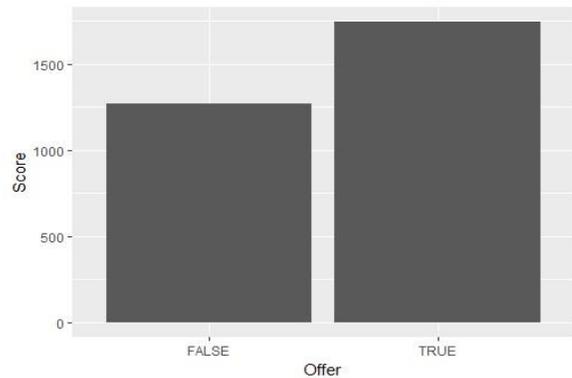


Fig. 3. Average Herbert scores of candidates.

Though candidate selection's correlation was significant, this is not a surprise. Fast Track uses a multi-method approach, such as using tools like an IQ test, whiteboard test, technical interviews, and manager interview. Ultimately the manager makes the decision based on all these factors. The strong correlation in candidate selection shows that in general, Herbert scores are in line with other indicators and with the hiring manager's final judgment.

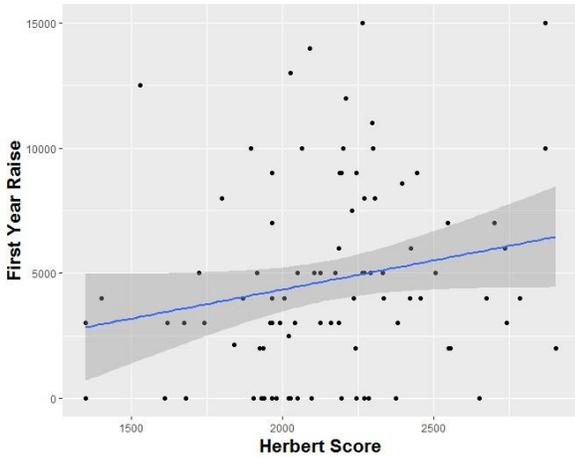
B. First Year Raise

Here the study correlated the candidate's Herbert score against their salary adjustment (raise) after their second review period (approximately one year).

This is a relevant metric because according to Conte, "candidates are hired at or close to the same compensation level. The first 3-6 months is often a trial period with little raises." However, Conte further states that "by the second review, managers are instructed to award raises based on proven value to date." Thus, the size of this raise is a good measure of the employee's value to the company in approximately the first year of their tenure.

For this correlation, the study considered just those 93 employees that stayed at least through the first year.

The study found a correlation of 19.3% with a P-Value of 0.03. This strongly suggests a relation between Herbert Score and an employee's first-year value to the company (Fig. 4).



4. Herbert score correlated with first year raise.

Fig.

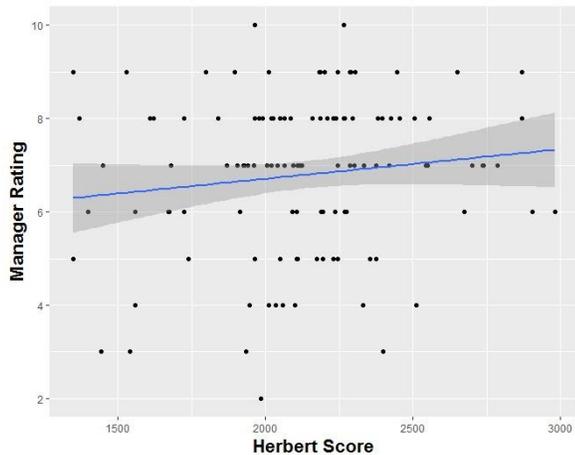


Fig. 5. Herbert score correlated with Manager's rating.

C. Manager's Rating Here the study correlated the candidate's Herbert score against the manager's rating at time of company exit. This reflects the aggregated value of the employee throughout their tenure, as estimated by the manager.

The average tenure of the employees in the Fast Track data set was 707 days, or just under two years, with a maximum of 3350 days, or just over 9 years.

The study found a correlation of 12.7% with a P-Value of 0.15. This suggests a possible relation between Herbert Score and long-term value to the company, as appraised by the manager (Fig. 5).

One reason this correlation might be weaker than the "first year raise," which should also be based on manager assessment, is that an employee's value may change over time due to many factors both inside and outside the company.

V. DISCUSSION

At Fast Track, the Herbert score is one of several factors used in deciding whether to offer a candidate a job. The use of the Herbert test at Fast Track appears to address the needs of tech companies' looking to find candidates who can solve

algorithmically challenging problems [13]. Not only does the Herbert test, handle complexity [14], it also addresses both convergent and divergent problem-solving [15]. There are specific solutions to the puzzles but the optimization goal, allows participants to engage in divergent problem solving. Interestingly problem-solving is associated with the reasoning portion of cognitive tests and intelligence has been linked to software developer success [9], [16].

Our results also suggest a link between a candidate's score in Herbert and their ultimate job performance. Thus, a test like Herbert appears to be a valid way to select valuable employees.

In conclusion, this study reinforces the need to develop tools like the Herbert test to help companies find skilled software developers. Both this study and previous literature support the significance of problem-solving as an important factor in the hiring process today and in the future. As technical systems further develop and demand for high skilled labor increases, "people who know how to use computers and intelligent algorithms optimally for their work will have the biggest opportunities in the job market" [18].

VI. FUTURE RESEARCH CONSIDERATIONS

This study reinforces literature that intelligence appears to be an important factor when making hiring decisions [9]. Yet, this is only one of several key considerations such as personality and knowledge of discipline [8], [13], [14]. Though this study does not directly look at the relationship between the Herbert test and cognitive skills, further research of this potential relationship should be considered in future analysis.

REFERENCES

- [1] Bureau of Labor Statitics, "Occupational Handbook Outlook," United States Department of Labor, 17 December 2015.
- [2] Winter, T., "True cost of recruiting a developer, Devskiller, 20 December 2016.
- [3] Dodge, J., "The war for tech talent escalates," Boston Globe, 19 February 2016.
- [4] Murphy, C. "What is the mean cost per year for a software engineer in silicon valley?" Quora, 2 March 2012.
- [5] Tech & Work, Tech companies have the highest turnover rate, TechRepublic, retrieved February 27, 2017. [6] Chakrabarty, R, "Only 7 percent engineering graduates employable: What's wrong with India's engineers?" India Today in Education, 13 July 2016.
- [7] Newdirectionsstaffing, "What a bad software developer hire could cost your company," Newdirectionsstaffing, 16 November 2015.
- [8] Google, "Life as a Software Engineer: Hiring process," Google, retrieved 27 February 2017.
- [9] Menkes, J., "Hiring for smarts," Harvard Business Review, November 2005.
- [10] Criteria, "Pre-Employment tests for computer software engineers," Criteria, retrieved 27 February 2017.
- [11] Wildnoodle, "Tutorial," Wildnoodle, retrieved 27 February 2017. [12] Interview Brian Conte, Founder and CEO FastTrack Team
- [13] Hewner and Guzdial (2010), "What game developers look for in a graduate: interviews and surveys at one game company," in ACM Digital Library, March 2010.
- [14] Li, L.P., Ko, A.J., and Zhu, J., "What makes a great software engineer?" in IEEE Press, Vol. 1, 2015, 700-710.

- [15] Graziotin, D., Xiaofeng, W. and Abrahamsson, P., "Happy software developers solve problems better: psychological measurements in empirical software engineering," in PeerJ, 11 March 2014.
- [16] Sundem, G., "Study: Is complex problem solving distinct from I.Q.?" in Psychology Today, 10 June 2013.
- [17] Liu, O.L., Frankel, L., and Roehr, K.C.. "Assessing critical thinking in higher education: Current state and directions for next-generation assessment," in ETS Research Report Series, 2 June 2014.
- [18] Braun, A., Zweck, A. and Holtmannspotter, D. "The ambiguity of intelligent: job killer or supporting assistant," in European Journal of Futures Research, 30 October 2016.