ABSTRACT

PALMIERI, DAVID WALSH. Knowledge Management Through Pair Programming. (Under the direction of Laurie Ann Williams.)

Knowledge Management has been the subject of increasing focus over the last several years. Literature and research on the topic has grown as companies and organizations have come to realize that success is often determined by one’s ability to create, disseminate, and embody knowledge in products and services. This realization has led to increased interest in examining the ways in which knowledge can be effectively created, identified, codified, disseminated, and retained. The field of Knowledge Management has emerged to address this need.

One of the obstacles that Knowledge Management seeks to overcome is the natural tendency in people to hoard knowledge. People often withhold knowledge when they feel it provides them with a competitive advantage over others. Many traditional management incentives and team structures create and perpetuate competitive environments that encourage knowledge hoarding. Knowledge Management also seeks to find ways to reduce the impact of employee turnover. When an employee leaves a company or organization, the knowledge they possess often goes with them. This loss can potentially have a negative impact on the productivity and quality of the company or organization. Knowledge Management seeks to find ways to minimize loss of knowledge when an employee leaves a company or organization.

Pair programming is a practice that holds promise for overcoming some of the challenges faced by Knowledge Management. In pair programming, two programmers
work side-by-side at one computer collaborating on the same design, algorithm, code, or test. The continual interaction between pair programmers would seem to provide an environment that promotes knowledge sharing, and collaborative knowledge discovery. Additionally, through pair rotation, in which pairs change partners fairly often, tacit knowledge might be spread more effectively through face-to-face communication than by documentation, databases, or other means.

This research examines pair programming in the realm of Knowledge Management, positing the following hypotheses:

- Pair programming reduces the tendency of people to hoard knowledge.
- Pair programming reduces the impact of employee turnover.
- Pair programming is an effective means of knowledge dissemination and knowledge retention that has a positive influence on the Knowledge Management practices of a company or organization.

These hypotheses are tested through the use of a survey of individuals in technology research, development, and service. Analysis of the survey results provided no conclusive evidence to either support or disprove the hypothesis that pair programming reduces the tendency of people to hoard knowledge. The results indicate support for the hypothesis that pair programming reduces the impact of employee turnover, although not statistically significant. And finally, the survey results indicate with statistical significance that pair programming is an effective means of knowledge dissemination and retention, with a positive influence on the Knowledge Management practices of a company or organization.
KNOWLEDGE MANAGEMENT THROUGH PAIR PROGRAMMING

by

DAVID WALSH PALMIERI

A thesis submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Master of Science

COMPUTER SCIENCE

Raleigh

2002

APPROVED BY:

_________________________________________________________

Chair of Advisory Committee
BIOGRAPHY

David Walsh Palmieri received a Bachelor of Science degree in Computer Science and a Bachelor of Arts degree in Psychology from Purdue University in 1985. After graduating, he joined the International Business Machines (IBM) Corporation as a software developer for their Systems/390 mainframe division in Poughkeepsie, New York. During ten years in the division, he designed and developed software for the company’s mainframe operating system and related products. In 1995, he moved to the Raleigh, North Carolina area, continuing to work for IBM at their Research Triangle Park facility, developing software for various workstation and security products. While working full time, he has also maintained an academic interest, pursuing a Master of Science degree in Computer Science at North Carolina State University. His more than fifteen years of software development experience in industry have enabled him to observe and participate in many processes and procedures used for developing commercial software.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2. BACKGROUND</td>
<td>4</td>
</tr>
<tr>
<td>2.1. Knowledge</td>
<td>4</td>
</tr>
<tr>
<td>2.2. Knowledge Management</td>
<td>8</td>
</tr>
<tr>
<td>2.2.1. Knowledge Management Models</td>
<td>8</td>
</tr>
<tr>
<td>2.2.2. Knowledge Management Elements</td>
<td>11</td>
</tr>
<tr>
<td>2.2.3. Knowledge Management Challenges</td>
<td>12</td>
</tr>
<tr>
<td>2.3. Pair Programming</td>
<td>14</td>
</tr>
<tr>
<td>2.3.1. Prior Research in Pair Programming</td>
<td>16</td>
</tr>
<tr>
<td>3. RESEARCH APPROACH</td>
<td>20</td>
</tr>
<tr>
<td>3.1. Survey Content and Format</td>
<td>20</td>
</tr>
<tr>
<td>3.2. Split Ballot</td>
<td>29</td>
</tr>
<tr>
<td>3.3. Pretesting</td>
<td>29</td>
</tr>
<tr>
<td>3.4. Sample Population and Distribution</td>
<td>30</td>
</tr>
<tr>
<td>4. RESULTS</td>
<td>32</td>
</tr>
<tr>
<td>4.1. Data Receipt, Recording, and Integrity</td>
<td>32</td>
</tr>
<tr>
<td>4.2. Data Analysis</td>
<td>33</td>
</tr>
<tr>
<td>4.3. Response Data</td>
<td>34</td>
</tr>
<tr>
<td>4.4. Amount of Pair Programming</td>
<td>34</td>
</tr>
<tr>
<td>5. ANALYSIS AND DISCUSSION</td>
<td>36</td>
</tr>
<tr>
<td>5.1. Pair Programming and Knowledge Management Effectiveness</td>
<td>36</td>
</tr>
<tr>
<td>5.1.1. Knowledge Source Penetration, Usage, and Ranking</td>
<td>36</td>
</tr>
<tr>
<td>5.1.2. Knowledge Foundation</td>
<td>42</td>
</tr>
<tr>
<td>5.1.3. Knowledge Management Effectiveness</td>
<td>44</td>
</tr>
<tr>
<td>5.2. Pair Programming and Knowledge Hoarding</td>
<td>51</td>
</tr>
<tr>
<td>5.3. Pair Programming and Employee Turnover</td>
<td>55</td>
</tr>
<tr>
<td>5.4. Subjective Responses</td>
<td>58</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 2.1. Williams et al: Percentage of Test Cases Passed ........................................... 18
Table 4.1. Likert Scale Values ......................................................................................... 33
Table 5.1. Knowledge Source Usage Responses and Values ........................................... 39
Table 5.2. Knowledge Management Effectiveness Responses and Values ................. 46
Table 5.3. Employee Turnover Responses and Values ...................................................... 56
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Knowledge Conversion</td>
<td>6</td>
</tr>
<tr>
<td>2.2</td>
<td>Knowledge Hierarchy</td>
<td>7</td>
</tr>
<tr>
<td>2.3</td>
<td>Pair Programming Workspace Layout</td>
<td>16</td>
</tr>
<tr>
<td>2.4</td>
<td>Williams et al: Relative Project Completion Times</td>
<td>18</td>
</tr>
<tr>
<td>4.1</td>
<td>Amount of the Day Spent Pair Programming</td>
<td>34</td>
</tr>
<tr>
<td>5.1</td>
<td>Percentage of Companies and Organizations Using Specific Knowledge Sources</td>
<td>38</td>
</tr>
<tr>
<td>5.2</td>
<td>Mean Usage of Knowledge Sources</td>
<td>39</td>
</tr>
<tr>
<td>5.3</td>
<td>Mean Ranking of Knowledge Sources</td>
<td>41</td>
</tr>
<tr>
<td>5.4</td>
<td>Mean Ranking of Knowledge Foundation Sources</td>
<td>43</td>
</tr>
<tr>
<td>5.5</td>
<td>Knowledge Management Effectiveness Index</td>
<td>47</td>
</tr>
<tr>
<td>5.6</td>
<td>Knowledge Dissemination and Pair Programming</td>
<td>49</td>
</tr>
<tr>
<td>5.7</td>
<td>Inversion of Room for Improvement</td>
<td>50</td>
</tr>
<tr>
<td>5.8</td>
<td>Knowledge as a Competitive Advantage</td>
<td>53</td>
</tr>
<tr>
<td>5.9</td>
<td>Willingness to Share Knowledge</td>
<td>54</td>
</tr>
<tr>
<td>5.10</td>
<td>Employee Turnover Impact Index</td>
<td>57</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

In today’s economy, with its shifting markets, tenacious competitors, and quickly-evolving technologies, companies and organizations are discovering that success is often determined by one’s ability to create, disseminate, and embody knowledge in products and services. Knowledge is rapidly displacing capital, financial and natural resources, and labor as the most important economic resource (Hansen et al, 1999). Forty percent of the U.S. economy is directly attributable to the creation of knowledge (Klasson, 1999).

Given that it holds such a financial stake, companies and organizations are interested in managing their knowledge effectively. Knowledge Management addresses this need by integrating human and organizational aspects with technological solutions. By applying Knowledge Management techniques, companies and organizations can improve their ability to create, acquire, disseminate, and retain knowledge, thereby enabling them to make effective decisions, control complexity, and improve productivity (Tiwana, 2000).

While the problems Knowledge Management strives to solve are numerous and far-reaching, two are of particular interest to this research: knowledge dissemination and knowledge retention. It has been noted that people have a natural tendency to keep knowledge to themselves, in the belief that it gives them power or competitive advantage over others (Liebowitz, 1999, and Probst et al, 2000). This makes it difficult to disseminate knowledge throughout a company or organization, dulling the effectiveness of the knowledge. Company and organizational knowledge assets are further impacted when an employee leaves, taking his or her personal knowledge with him/her.
Companies and organizations seek for ways to guard against such loss, given the importance of knowledge in today’s economy.

Pair programming could provide an aid to these challenges. Research has indicated that pair programming is an effective software development practice that produces higher quality code in roughly the same amount of effort as traditional individual programming (Nosek, 1998, and Williams et al, 2000). A potential benefit of pair programming that has not yet been investigated is its effectiveness as a Knowledge Management technique. According to Probst et al, “knowledge is a commodity which is often only transferable in personal exchanges between individuals” (Probst et al, 2000). The close, prolonged personal communication inherent in pair programming would seem likely to reduce knowledge hoarding and facilitate dissemination. Additionally, pair rotation, in which pairs change partners fairly often, would seem likely to reduce the impact of employee turnover.

This research combines two relatively recent fields of study – Knowledge Management and pair programming – to see if one (pair programming) has a positive effect on the other (Knowledge Management). Specifically, it investigates the following hypotheses:

- Pair programming reduces the tendency that people have to hoard knowledge.
- Pair programming reduces the impact of employee turnover.
- Pair programming is an effective means of knowledge dissemination and knowledge retention that has a positive influence on the Knowledge Management practices of a company or organization.
These hypotheses are evaluated based on input from a survey of individuals in technology research and industry.

The work is laid out in the remaining chapters of this paper:

- Chapter 2. BACKGROUND contains a brief investigation of the concepts upon which this research is based. These concepts include knowledge – its categories and hierarchies, Knowledge Management, and pair programming.
- Chapter 3. RESEARCH APPROACH describes the method used by this work to assess the effectiveness of pair programming as a Knowledge Management technique. It describes the creation of a survey that was conducted to test the hypotheses associated with this research.
- Chapter 4. RESULTS presents a summary of the data produced by the survey related to this work.
- Chapter 5. ANALYSIS AND DISCUSSION evaluates the data to provide a meaningful context and draw inferences.
- Chapter 6. CONCLUSIONS contains a summary of the work, and the major findings associated with this study.
- Chapter 7. LIST OF REFERENCES enumerates the prior work that was studied and incorporated into this research.
- Chapter 8. APPENDICES includes a copy of the survey used in this research, as well as additional detailed information regarding the procedure and data associated with this study.
2. BACKGROUND

A comprehensive investigation of any topic requires examination of the issues surrounding it. This section provides an overview of knowledge, including its categories and hierarchies, a review of Knowledge Management’s processes, and the mechanics of pair programming. The background information presented here serves as a basis, referred to and built upon by the investigation as it goes to greater depths.

2.1. Knowledge

Most everyone has an intuitive idea of what knowledge is. Webster’s dictionary defines knowledge as follows (Webster’s, 1993):

The fact or condition of knowing something with a considerable degree of familiarity gained through experience of or contact or association with the individual or thing so known.

Perhaps more appropriate to the subject at hand is the definition put forth by Davenport and Pruzak (Davenport and Pruzak, 1998):

Knowledge is a fluid mix of framed experience, values, contextual information, expert insight and grounded intuition that provides an environment and framework for evaluating and incorporating new experiences and information. It originates and is applied in the minds of knowers. In organizations, it often becomes embedded not only in documents or repositories but also in organizational routines, processes, practices, and norms.

At a broad level, knowledge can be classified into two categories, tacit and explicit knowledge (Nonaka and Takeuchi, 1995). Tacit knowledge is personal, context-
specific knowledge that is difficult to formalize, record, or articulate; it is stored in the heads of people. Explicit knowledge is knowledge that can be codified and transmitted in a systematic and formal language, such as documents, databases, e-mails, etc.

Nonaka and Takeuchi claim that most knowledge is tacit, with only a small portion of it making its way to the explicit form through the process of externalization (Nonaka and Takeuchi, 1995). Externalization consists of recording knowledge into the form of documents, databases, tools, etc. Once tacit knowledge is externalized and converted into explicit knowledge, the knowledge can be converted back into tacit knowledge through the process of internalization. Internalization consists of applying training, processes, communication, and practice to make knowledge almost second nature to those requiring it. Most tacit knowledge is never externalized, but is transferred from one person to another via the process of socialization. Socialization consists of informal sharing of knowledge between two or more people. And finally, external knowledge can be converted to new external knowledge through the process of combination. Combination occurs when external knowledge is combined from two or more sources to create a richer knowledge.

These knowledge conversion processes are summarized in Figure 2.1 (Nonaka and Takeuchi, 1995):
Similarly, Le Pechoux and Honeycutt propose a model of knowledge transition along the dimensions of internal and external, discovery and refinement (Le Pechoux and Honeycutt, 2001). Internal knowledge consists of creative achievements developed internally, and external knowledge consists of knowledge in the external environment. Discovery involves the consolidation and synthesis of what has been learned and refined, and refinement represents solution increments. Knowledge can transition from the internal to external state through sensing, in which knowledge is recognized and documented. Knowledge can transition from the external state to the internal state via feeling, in which knowledge is sought and/or learned. Transitions along the discovery and refinement dimension are achieved through thinking and intuition. Knowledge can transition from discovery to refinement through thinking, in which logic and adjustments are applied to existing knowledge. Knowledge can transition from refinement to discovery via intuition, in which knowledge is perceived and/or organized in a new way.

Besides classification, knowledge can be organized along a hierarchical dimension. Beckman proposes the following knowledge hierarchy (Beckman, 1997):
The model identifies the following levels in the hierarchy:

1. Data, consisting of text, facts, images, and sound.

2. Information, consisting of organized, structured, interpreted, and summarized data.

3. Knowledge, in which reasoning, abstraction, and relationships are applied to information to produce cases, rules, processes, and models.

4. Expertise, in which selection, experience, and principles are applied to knowledge to produce fast and accurate advice, explanations, and justifications based on reasoning.

5. Capability, in which integration and distribution are applied to expertise to produce organizational knowledge repositories and core competencies.

The distinction between data, information, and knowledge is the level of organization and abstraction. As the organization and abstraction increases, so too does its value. According to Beckman, knowledge can be transformed from a lower level to a more valuable higher level by applying increasingly higher degrees of organization and abstraction.
Given that most knowledge is tacit, companies and organizations potentially have much to gain by tapping into the tacit knowledge that exists and finding ways to effectively disseminate it.

2.2. Knowledge Management

Different sources have different definitions of Knowledge Management. The term was coined by Wiig, who defines it as follows (Wiig, 1997):

*Knowledge Management is the systematic, explicit, and deliberate building, renewal, and application of knowledge to maximize an enterprise’s knowledge-related effectiveness and returns from its knowledge assets.*

Another definition that is consistent with other sources is that provided by the Gartner Group (from Morey et al, 2000):

*Knowledge Management is a discipline that promises an integrated approach to identifying, managing and sharing all of an enterprise’s information assets, including database, documents, policies and procedures as well as unarticulated expertise and experience resident in individual workers.*

Knowledge Management is a relatively new discipline, having evolved from the expert systems of the 1980s. Industry and research interest in Knowledge Management has increased recently as companies and organizations have realized the value of their knowledge assets.

2.2.1. Knowledge Management Models

Models for the Knowledge Management process aim to provide direction for effectively formalizing, distributing, sharing, and applying knowledge. Many different
Knowledge Management models exist. O’Dell proposes a model for managing knowledge that is concise, yet consistent with most others (O’Dell, 1996). The model identifies the following steps for synergistically managing knowledge:

1. Identify, in which the company or organization determines its core competencies, sourcing strategy, and knowledge domains
2. Collect, in which existing knowledge is formalized
3. Adapt, in which knowledge is adapted to meet the specific needs of the organization
4. Organize, in which knowledge is stored in a knowledge repository so that it can be easily located when required
5. Apply, in which knowledge is used to make decisions, solve problems, and improve processes
6. Share, in which knowledge is distributed to those who need it
7. Create, in which new knowledge is discovered through research, experimentation, and creative thinking

All of these activities are often concurrent, repeated, and not always in sequence. However, together they enable the company or organization to more effectively use their knowledge.

Le Pechoux and Honeycutt propose a model that is based firmly in the tradition of Patterns (Alexander, 1977; Le Pechoux and Honeycutt, 2001). The model, called the Knowledge Insight Model, defines four roles, the Framer, the Sharer, the Maker, and the Finder:
• The *Framer* focuses on planning the Knowledge Management activity by determining the Knowledge Management techniques required to achieve set objectives within the company or organization. The Framer is responsible for understanding the environment, identifying base parameters, and developing a plan.

• The *Sharer* focuses on tracking existing knowledge and integrating new knowledge. Activities of the Sharer include developing tools to implement control measurements, monitoring results to identify potential problems, and documenting impacts.

• The *Maker* focuses on creating new products, concepts, and processes, thereby stimulating knowledge creation. The Maker examines knowledge resident in the company or organization and aggregates it to create new knowledge required to meet objectives. The Maker also determines objectives for discovering new knowledge outside of the company or organization.

• The *Finder* focuses on optimizing knowledge deployment based on process and organizational needs. The Finder deploys new products, concepts, and processes, and identifies requirements for new ones.

Roles are not necessarily assigned to specific individuals. Sometimes multiple people assume the same role, and sometimes one person assumes multiple roles.

Regardless of how the roles are assigned, the level of interaction between roles often determines the model’s effectiveness.
2.2.2. Knowledge Management Elements

Many Knowledge Management models include elements of technology, organization, and management. Tobin suggests an information technology infrastructure consisting of (Tobin, 1998):

- Knowledge repository, most commonly a database, that organizes and makes available to all employees basic information on the company’s organization, products, services, customers, and business processes
- Directory of knowledge sources that can be used by employees to locate specific knowledge when needed
- Directory of learning sources that can be used by employees to find out where they may learn new knowledge and skills
- Groupware that facilitates communication and collaboration among employees

Other information technology elements common to Knowledge Management models include expert systems, decision support systems, and integrated performance support systems.

For the organizational element of Knowledge Management, Davenport and Pruzak suggest the following roles (Davenport and Pruzak, 1998.):

- Knowledge-oriented personnel, which includes all people who create, share, search for, or use knowledge in their daily work
- Knowledge Management Specialist, responsible for representing, organizing, and maintaining content in the knowledge repository
• Knowledge Project Manager, whose activities include developing project objectives, assembling and managing teams, determining and managing customer expectations, monitoring project budgets and schedules, and identifying and resolving project problems

• Chief Knowledge Officer, who serves as an advocate or evangelist for knowledge and learning, designs, implements, and oversees the organization’s knowledge infrastructure, and is the primary liaison between external providers of information and knowledge

Management aspects of Knowledge Management focus on creating an environment that is receptive and committed to knowledge organization. Measuring performance, establishing a reward system based on those measurements, and encouraging knowledge exploration and sharing are all key elements.

2.2.3. Knowledge Management Challenges

While the scope of Knowledge Management is rather broad, two specific areas that many Knowledge Management strategies focus on are knowledge dissemination and knowledge retention. Knowledge dissemination is negatively impacted when people withhold knowledge in the belief that it gives them a competitive advantage over others (Beckman, 1999):

In bureaucratic organizations, employees and managers are discouraged from sharing knowledge and expertise. In fact, the opposite is often the case: knowledge is considered a source of power, and thus hoarding is not only expected but is often rewarded.
Beckman suggests overcoming the tendency to hoard knowledge by cultivating a management structure that encourages sharing and rewards teamwork. By doing so, individuals are more likely to share knowledge with others, allowing for greater dissemination.

Several sources also claim that knowledge is most effectively disseminated through face-to-face communication. According to Probst et al (Probst et al, 2000), “knowledge is a commodity which is often only transferable in personal exchanges between individuals.” This viewpoint is supported by Wathne et al (Wathne et al, 1996):

*Our framework proposes an indirect causal relationship between the richness of the channel of interaction and the effectiveness of knowledge transfer. We argue that face-to-face interaction is the richest medium because of its capacity for immediate feedback…and the availability of multiple cues…It creates the richest, most open social context through which knowledge is transferred.*

Knowledge dissemination can also aid in knowledge retention when employees leave an organization. Knowledge Management strategies should focus on capturing an individual’s knowledge while they are in the organization, so that the knowledge is not lost when the individual leaves the company or organization. Given that its knowledge assets increasingly determine the value of a company or organization, it is becoming more and more critical that organizations guard against such loss. By disseminating each and every individual’s knowledge throughout an organization, knowledge loss due to employee turnover is minimized.
During its rather brief history, Knowledge Management has quickly evolved into a discipline of great interest to many companies and organizations. By focusing on knowledge creation, dissemination, and retention, Knowledge Management identifies strategies for managing knowledge assets. With the realization of the value of knowledge assets, companies and organizations increasingly seek to implement Knowledge Management and increase its effectiveness.

2.3. Pair Programming

Before commencing with an assessment of its effectiveness as a Knowledge Management technique, pair programming will be described in greater detail.

As mentioned previously, pair programming consists of two programmers working side-by-side at one computer, collaborating on the same design, algorithm, code, or test. One programmer is the driver, controlling the input device (keyboard and mouse) to produce the design or code. The other programmer is the navigator, continuously and actively examining the driver’s work. The navigator watches for defects, thinks of alternatives, looks up resources, considers strategic implications, and asks questions. Through these activities, the navigator contributes by identifying tactical and strategic deficiencies in the design or code.

Every so often, the driver and navigator switch roles. This allows the programmers to vary their work routine, learn and practice new skills, and contribute to the work in more than one manner. There is no set rule as to how often the programmers should switch roles. It should occur at a natural transition point in the design or coding activity. Sometimes this may be after only one or two hours, or sometimes it may be
after days. In organizations where there is more than one pair, pairs often rotate partners after the completion of a task or subtask. Pair rotation allows the programmers to learn different areas of the system, and learn from the skills of different partners.

When two programmers are first paired, there is often a jelling period during which time each is adjusting to the working habits of the other, learning the other’s strengths and weaknesses, and learning how to communicate effectively. While productivity is somewhat less during this period, it soon increases as the pair becomes comfortable working together. Anecdotal evidence suggests that this adjustment period varies from a few hours to several days, depending on the individuals (Williams et al, 2000).

There are variations in the way that organizations structure their workplace environment to support pair programming. The most common configuration is to use one monitor, one keyboard, and one mouse. The programmers sit next to each other so that they each have a clear view of the monitor. Control can be passed by simply sliding the keyboard and mouse, or less optimally by sliding chairs. Other less common configurations are two monitors, one keyboard, one mouse; one monitor, two keyboards, and two mice; or two monitors, two keyboards, and two mice. In order to facilitate sitting next to each other and sliding control, pair programmers often situate their chairs and equipment around open, rather than closed, corners, as illustrated in Figure 2.3 (from Williams et al, 2000):
The pair programming approach lends itself to several potential benefits. By being under the scrutiny of two programmers at all times, the design and code is under continuous review, potentially leading to higher defect-removal rates. The combined strengths of two programmers can potentially enable them to solve difficult problems more quickly than if they had worked alone. By pairing, programmers can learn skills from each other and communicate more closely. And, exposing multiple programmers to each piece of code potentially reduces the impact of losing staff.

2.3.1. Prior Research in Pair Programming

The history of formal research in pair programming goes only as far back as 1998. In that year, Nosek carried out an experiment in which he studied 15 experienced programmers working up to 45 minutes on a challenging problem related to their work (Nosek, 1998). Five programmers worked individually, and 10 programmers worked in pairs. The conditions and material were the same for the individuals as the pairs. Combining their time, the pairs spent 60% more time on the task, but because they worked together, they completed the task in 40% less time than the individuals. In
addition to requiring less time, the pairs produced better algorithms and code than the individuals.

Nosek’s study was somewhat narrow in scope, drawing conclusions from only 45 minutes of pair programming time. In 1999, Williams et al, carried out an experiment of considerably broader scope, studying 41 senior software engineering students over the span of four programming assignments (Williams et al, 2000). The students were split into two groups, both composed of the same mix of high, medium, and low performers based on grade-point average. The first group, consisting of 13 students, worked individually on all assignments. The second group, consisting of 28 students, worked as pairs (14 pairs in all) on all of the assignments. To make the workload even, the pairs were given additional assignments above and beyond what was assigned to the individuals.

During the first assignment, a jelling period was evident, as the pairs took 60% more time than the individuals to complete the assignment. By the completion of the second assignment, the jelling period had ended, reflected by the fact that the pairs required only 15% more time than the individuals to complete the assignment. Figure 2.4 shows the amount of time required by the pairs and the individuals to complete the first three programming assignments:
Although the pairs were slightly less productive than the individuals, the study found that the pairs produced higher quality code than the individuals. The code developed by the pairs passed 15% more of the teaching staff’s automated test cases than the code developed by the individuals. The percentage of test cases passed by the individually programmed code and the pair programmed code are shown in Table 2.1:

Table 2.1. Williams et al: Percentage of Test Cases Passed

<table>
<thead>
<tr>
<th></th>
<th>Individuals</th>
<th>Pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program 1</td>
<td>73.4</td>
<td>86.4</td>
</tr>
<tr>
<td>Program 2</td>
<td>78.1</td>
<td>88.6</td>
</tr>
<tr>
<td>Program 3</td>
<td>70.4</td>
<td>87.1</td>
</tr>
<tr>
<td>Program 4</td>
<td>78.1</td>
<td>94.4</td>
</tr>
</tbody>
</table>
The results of the experiment were quite encouraging, indicating that pair programming could produce higher quality code in relatively the same amount of time as individual programming.

One other aspect of pair programming that has been the focus of research is the level of satisfaction associated with it. Cockburn and Williams conducted an anonymous survey of professional pair programmers and student pair programmers to determine whether or not pair programming was more enjoyable than individual programming (Cockburn and Williams, 2000). The results of the survey showed that more than 90% of those that practice pair programming enjoy their work more because of it.

The prior research in pair programming indicates that it produces higher quality code than that produced by traditional individual programming, in roughly the same effort. However, no research has been performed thus far to assess the effectiveness of pair programming as a Knowledge Management technique.
3. RESEARCH APPROACH

In order to test the hypotheses concerning pair programming as a Knowledge Management technique, a survey was created and given to a cross sectional sample of programmers in information technology research and industry. The survey was designed to examine the differences between companies and organizations that practice pair programming, and those that do not.

A survey was chosen as the research method because it allowed a geographically diverse population to be sampled relatively quickly and at a low cost. Personal interviews were ruled out by cost and time, and considered less convenient for participants. Several methods of distribution were considered, including paper, electronic document, e-mail, and web-based. In the end, a text-based e-mail format was chosen because it allowed for fast distribution, required no supporting server infrastructure, and was convenient for participants.

The survey is included in the appendix section of this paper.

3.1. Survey Content and Format

The format of the survey consisted of an introduction explaining the source and intent, general instructions, and a series of questions. Most questions were closed-ended, consisting of a statement or question, followed by a multiple of choices from which an answer could be chosen. Closed-ended questions were used because they provide the same frame of reference for all participants to use when determining answers, and it is easy and inexpensive to work with the resulting data (Weisberg et al, 1996). For some questions, where it was particularly difficult to anticipate all possible choices,
participants were given the opportunity to specify their own answer if it was not among
the list of choices provided. Each choice associated with a closed-ended question was
prefixed with an empty box (“[ ]”) in which the participant could indicate his/her
selection by placing a non-blank character inside the box. A few questions asked the
participant to rank possible choices, in which case he/she was required to put a number
inside the empty box.

The survey also included some open-ended questions designed to elicit subjective
responses. As noted by Bainbridge, while somewhat more difficult to work with, open-
ended questions are “the best way to gain complex, freely expressed ideas, impressions,
and personal experiences.” (Bainbridge, 1989)

In all, the survey included 29 questions. Questions were divided into three
sections:

- Section 1: Knowledge Sources
- Section 2: Knowledge Acquisition, Dissemination, and Maintenance
- Section 3: General

Section 1 (Knowledge Sources) consisted of 10 questions designed to test the
hypothesis that pair programming is an effective means of knowledge dissemination and
knowledge retention that has a positive influence on the Knowledge Management
practices of a company or organization. So as not to overload the participant with
Knowledge Management terminology, these questions were posed from the perspective
of investigating the sources utilized (e.g., books, co-workers, databases, etc.) when
searching for knowledge to get work done. One of the questions and some commentary follow:

*When a person requires some additional knowledge or information to get their job done, they sometimes turn to various sources for the information they are seeking. When you are faced with that situation, how often do you find the information you are looking for in the following sources:*

1. *Internal publications (such as guidelines or “How-To” documents, not design documents) produced by your company or organization.*
   - My company or organization does not produce internal publications of this sort.
   - Often.
   - Sometimes.
   - Rarely or never.

The intent of these questions was twofold:

1. Identify the Knowledge Management techniques employed by the participant’s company or organization

2. Assess the effectiveness of each technique by determining how often it is utilized

In addition to indicating how often they used each technique, participants were asked to rank the techniques in the order in which they were most likely to be used:
9. **Please rank the following in the order you are MOST LIKELY to turn when seeking additional information to get your job done. Place a 1 next to the source you are most likely to go to first, a 2 next to the source you are most likely to go to second, and so on.**

[ ] Internal publications (such as guidelines or “How-To” documents, not design documents) produced by your company or organization.

[ ] …etc.

Participants were also asked to rank the order of the sources that contributed most to learning their job:

10. **Please rank the following in the order in which you obtained the most knowledge to get your current, day-to-day job done**

[ ] Academic education (e.g., college or high school).

[ ] …etc.

Section 2 (Knowledge Acquisition, Dissemination, and Maintenance) consisted of 11 questions designed to test all of the hypotheses. Seven questions continued testing the hypothesis that pair programming is an effective means of knowledge dissemination and knowledge retention that has a positive influence on the Knowledge Management practices of a company or organization. One of the seven questions asked the participant to indicate whether or not their company or organization had a Knowledge Management organizational strategy in place:
15. *Does your company or organization have personnel in place specifically responsible for managing knowledge* (e.g., *Chief Knowledge Officer, Knowledge Project Manager, Knowledge Management Specialist, Knowledge Team, etc.*)?

[ ] Yes.

[ ] No, or not that I am aware of.

Question 19 asked the participant to indicate whether or not their company or organization had a management strategy in place that supported Knowledge Management:

19. *In your company or organization, how much are rewards based on individual technical accomplishments, versus team success?*

[ ] Virtually all individual accomplishments.

[ ] Primarily individual accomplishments, but also some team success.

[ ] Equally individual accomplishments and team success.

[ ] Primarily team success, but also some individual accomplishments.

[ ] Virtually all team success.

The other five of the seven questions asked the participant to assess the overall effectiveness of the Knowledge Management strategies employed by their company or organization. For four of these five questions, choices were presented using a Likert scale running from “strongly agree” to “strongly disagree.” (Likert, 1932):

11. *My company or organization is good at creating new knowledge through its people and technological resources.*
12. My company or organization is good at finding, organizing, and documenting the knowledge it already possesses.
[ ] …etc.

13. My company or organization is effective at acquiring knowledge from outside sources, such as consultants or products.
[ ] …etc.

14. My company or organization is effective at making knowledge accessible to those who need it, when they need it.
[ ] …etc.

The last of the Knowledge Management strategy effectiveness questions did not use a Likert scale:

16. How much improvement in the quality and productivity of your work could be gained by improvements to your company’s or organization’s management of knowledge:
Besides the seven questions designed to test the hypothesis regarding the effectiveness of pair programming as a Knowledge Management technique, Section 2 included two questions designed to test the hypothesis that pair programming reduces the tendency that people have to hoard knowledge:

17.  I feel the unique knowledge I possess enhances my competitive advantage over my peers when it comes to job promotions, leadership opportunities, and awards.

[ ] Strongly agree.
[ ] Agree.
[ ] Neither agree nor disagree.
[ ] Disagree.
[ ] Strongly disagree.

18.  I am willing to share all of the knowledge I possess relevant to my job with my company or organization

[ ] ...etc.

And finally, Section 2 included two questions designed to test the hypothesis that pair programming reduces the impact of employee turnover:
20. *If you were to leave your position tomorrow, how long would it take your organization to compensate for the loss of knowledge you possess? (Please make your best guess)*

[ ] Less than one week.

[ ] Between one week and one month.

[ ] Between one month and three months.

[ ] Between three months and six months.

[ ] Between six months and one year.

[ ] More than one year.

21. *If the most senior technical member of your organization were to leave his/her position tomorrow, how long would it take your organization to compensate for the loss of knowledge he/she possesses? (Please make your best guess)*

[ ] ...etc.

Section 3 (General) consisted of eight questions designed to gather demographical background information and subjective information. These questions were not designed to test a specific hypothesis, but rather obtain information related to all of the hypotheses. To test any of the hypotheses, it was necessary to distinguish pair programmers from non-pair programmers. For this purpose, one question was included to assess the degree to which the participant practiced pair programming:
27. *Pair programming* is a practice in which two programmers work side-by-side at one computer, collaborating on the same design, algorithm, or test. Approximately how much of your day is spent pair programming?
   
   [ ] Less than 10%.
   
   [ ] Between 10% and 25%.
   
   [ ] Between 25% and 50%.
   
   [ ] Between 50% and 75%.
   
   [ ] More than 75%.

In addition, one question was included to assess the degree to which the participant collaborated with others, regardless of whether or not they pair programmed:

26. Approximately how much of your day is spent collaborating with others?
   
   [ ] ...etc.

Four questions gathered basic demographic information:

22. How long have you been a programmer in industry or research?

23. How long have you been with your current employer?

24. How satisfied are you with your job?

25. What software methodologies do you use in your current organization?

   Please indicate all that apply.

Finally, two open-ended questions asked participants to discuss the strengths and weaknesses of the Knowledge Management strategies of their company or organization:

28. *(Optional)* What do you consider to be the strengths of your company’s or organization’s management of knowledge?
29. *(Optional) What do you consider to be the weaknesses of your company’s or organization’s management of knowledge?*

### 3.2. Split Ballot

To reduce the possibility of bias due to question and/or answer sequencing, a technique called “split-ballot” (or “split-sample”) was used (Leithold, 1982). In a split-ballot, multiple versions of a survey are prepared, each with a different order of questions and/or answers. The variation in order reduces the possibility of bias due to ordering. For this research, three different versions of the survey were prepared. For all three versions, section ordering remained constant. That is, Section 1 (Knowledge Sources) was presented first, Section 2 (Knowledge Acquisition, Dissemination, and Maintenance) was presented second, and Section 3 (General) was presented last. Section 3 (General) was specifically placed at the end of the survey so that the demographic and open-ended questions could be presented after having built a rapport with the participant through the preceding questions, as recommended by Burgess (Burgess, 1976).

In the final distribution of the survey, an equal number of each version was distributed to the sample population.

### 3.3. Pretesting

As noted by Fowler, pretesting a survey by giving it to a small number of participants prior to larger distribution has specific benefits (Fowler, 1988):

*Every questionnaire should be pretested, no matter how skilled the researcher.*

*Virtually every questionnaire could be changed in some way to make it easier for*
respondents and interviewers to meet the researcher’s objectives... One outcome of a good pretest is to find out how long it takes to complete a questionnaire.

Prior to general distribution, a pretest was performed on a small subset of the sample population. Pretest participants were asked to keep track of and report how long it took them to complete the survey. Responses from the pretest were examined to identify any problems with the questions and/or answers that might cause confusion. As a result of the pretest, slight modifications were made to the survey before it was distributed to the sample population.

3.4. Sample Population and Distribution

Because this research required a comparison of the Knowledge Management practices of pair programmers and non-pair programmers, it was vital that the cross sectional sample population consist of both pair programmers and non-pair programmers. Potential participants were identified from the following sources:

- Programmers who had previously participated in a pair programming survey conducted by Williams and Kessler to gather data for a book on pair programming (Williams and Kessler, 2002). It was thought that this would be a good source of pair programmers.
- Programmers who had recently visited pairprogramming.com, another likely source of pair programmers.
- Members of the extremeprogramming Yahoo group. Given that pair programming is one of the 12 practices of the Extreme Programming software
development process, it was thought that this would be a likely source of pair programmers (Beck, 2000).

- Members of other Yahoo groups, including boost (C++ related) and runrtp (athletic related, consisting primarily of individuals in technology research and industry). It was thought that this would be a likely source of non-pair programmers.

- Personal contacts in industry.

Drawing upon these sources, the survey was sent to over 500 individuals. To increase the likelihood of getting responses, a technique called “snowball sampling” was used, in which participants were encouraged to forward the survey to colleagues, particularly those in other organizations. Snowball sampling is frequently used when attempting to acquire evidence from a rare population (Weisberg et al, 1996).

Initial returns of the survey indicated a lack of participation from pair programmers. To increase the number of pair programmer participants, the survey was distributed to likely pair programmers a second time, encouraging them to participate, and it was posted publicly to the extremeprogramming group on Yahoo, which at the time consisted of over 3000 members.

In all, over 3500 people were contacted either individually or through the extremeprogramming group on Yahoo.
4. RESULTS

Following the distribution of the survey to the cross sectional sample of programmers in information technology research and industry, responses were received and recorded. Preliminary analysis was performed on the data to ensure integrity.

4.1. Data Receipt, Recording, and Integrity

Responses were received via e-mail. Each response was filed on two different computer systems, and printed out on paper to reduce the likelihood of losing data. The split-ballot distribution required that responses be sorted by version number (1, 2, or 3) before data was recorded. Data was then recorded into a Microsoft Excel worksheet, one response at a time, in the order in which they were received. Data entry was complicated by the fact that each version of the survey contained questions and answers in a different order. For the data entry, all responses were recorded using the format established by version 1. This required that questions and answers from versions 2 and 3 of the survey be transcribed to the format of version 1.

For all but the open-ended questions, responses were recorded as numbers. For example, for the questions using a Likert scale, responses were recorded using the following values:
Table 4.1. Likert Scale Values.

<table>
<thead>
<tr>
<th>Response</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly agree</td>
<td>2</td>
</tr>
<tr>
<td>Agree</td>
<td>1</td>
</tr>
<tr>
<td>Neither agree nor disagree</td>
<td>0</td>
</tr>
<tr>
<td>Disagree</td>
<td>-1</td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>-2</td>
</tr>
</tbody>
</table>

After all of the responses were recorded into the Microsoft Excel worksheet, some preliminary analysis was done to detect data entry errors. To further ensure data integrity, all responses in the Microsoft Excel worksheet were printed off on paper and re-entered into another Microsoft Excel worksheet working off of the printed paper. This uncovered several more errors that had not been detected by the preliminary analysis.

4.2. Data Analysis

Data analysis was performed using both Microsoft Excel, and Statistical Package for the Social Sciences (SPSS). After the preliminary analysis was done in Microsoft Excel to uncover recording errors, further analysis was performed using Microsoft Excel to produce frequency distributions and mean values for responses. More advanced statistical analysis required for question aggregation, response correlation, and statistical significance was performed using SPSS.
4.3. Response Data

In all, 95 responses were received. The survey consisted of 27 closed-ended questions, and two optional open-ended questions. Of the cumulative of 2565 closed-ended questions posed across all participants, 24 were left unanswered or answered invalidly (e.g., more than one answer specified when only one was allowed), representing 0.9%. Of the cumulative 190 optional open-ended questions posed across all participants, 112 (59%) were left unanswered.

4.4. Amount of Pair Programming

Figure 4.1 shows the distribution of responses regarding the amount of pair programming performed by participants:

![Bar chart showing the distribution of responses regarding the amount of pair programming performed by participants.](image)

Figure 4.1. Amount of the Day Spent Pair Programming

Even though extra effort was expended encouraging pair programmers to participate in the survey, far more responses were received from non-pair programmers.
than pair programmers. This is representative of the software industry; far more programmers work solo than paired. Although the low number of responses from pair programmers does not invalidate any analysis based on the amount of pair programming performed by the respondent, it does require a greater magnitude in the difference to reach statistical significance than if the numbers had been more evenly distributed.
5. ANALYSIS AND DISCUSSION

Using the responses from the survey of a cross sectional sample of programmers in industry and research, analysis was performed to test the hypotheses concerning pair programming as a Knowledge Management technique. This section presents the analysis of the data and discusses the results with regards to the three hypotheses of this research, as stated on page 2 of the INTRODUCTION section.

5.1. Pair Programming and Knowledge Management Effectiveness

The entire first section of the survey, and a portion of the second, contained questions designed to test the hypothesis that pair programming is an effective means of knowledge dissemination and knowledge retention that has a positive influence on the Knowledge Management practices of a company or organization.

5.1.1. Knowledge Source Penetration, Usage, and Ranking

The first nine questions of the survey examined various knowledge sources to determine how often and how effectively they are used. Participants were asked whether or not, and how often, they used the following knowledge sources:

- Internal publications, such as guidelines or “How-To” documents, not design documents, produced by the participant’s company or organization.
- Design documents, from either the project at hand or a previous project, produced by the participant’s company or organization.
- External publications, such as books, manuals, journals, or magazines.
- The Internet.
- Co-workers within the participant’s immediate team or organization.
• Courses, either classroom or online.

• A skills “Yellow Pages” that allows people to find others within the company or organization based on skill.

• A database or groupware, such as SAP or Lotus Notes.

For the purposes of this research, co-workers were considered to be the knowledge source most closely related to the practice of pair programming. The focus of the analysis for these questions was examine responses across all participants, regardless of the amount of pair programming performed. If it could be shown that knowledge sources most closely related to pair programming (i.e., co-workers) are generally more useful and effective than others, it would lend support to the hypothesis that pair programming is an effective means of knowledge dissemination and retention. While a breakdown of the responses by pair programming might be interesting, it would only serve to identify unique patterns in the way pair programmers use knowledge sources, which is beyond the scope of this research.

Figure 5.1 shows percentage respondents whose companies and organizations utilize each knowledge source:
External publications and the Internet were the only two knowledge sources that achieved 100% penetration, with co-workers and courses falling close behind at 99% and 98%, respectively. Far below the other knowledge sources was the skills “Yellow Pages,” used by only 59% of the companies and organizations participating in the survey.

Besides penetration, another interesting aspect of knowledge source utilization is effectiveness. Participants were asked to report how often they used each knowledge source successfully. Possible responses, and the corresponding values used for data analysis, are shown in Table 5.1:
Table 5.1. Knowledge Source Usage Responses and Values

<table>
<thead>
<tr>
<th>Response</th>
<th>Numerical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>My company or organization does not use this knowledge source</td>
<td>Not counted in the effectiveness analysis</td>
</tr>
<tr>
<td>Often</td>
<td>2</td>
</tr>
<tr>
<td>Sometimes</td>
<td>1</td>
</tr>
<tr>
<td>Rarely or never</td>
<td>0</td>
</tr>
</tbody>
</table>

Applying these metrics to the knowledge source usage responses across all respondents, the following mean usage values were calculated for each knowledge source:

![Knowledge Source Usage](image)

Figure 5.2. Mean Usage of Knowledge Sources
According to the results, the Internet, co-workers, and external publications were the three most effectively used knowledge sources. A one-sample T-test was used to determine whether or not the difference in mean usage across different knowledge sources was statistically significant (when statistical significance is defined as $p < .05$). The analysis revealed that the difference in mean usage was statistically significant between the Internet, co-workers, external publications, and design documents ($p < .05$). Design documents, internal publications, and databases were all statistically equivalent. The difference in mean usage between databases and courses was not statistically significant, but the difference between internal publications and courses was ($p < .05$). And finally, the mean usage for courses was significantly higher than that of the skills “Yellow Pages” ($p < .05$).

Besides reporting penetration and usage, participants were asked to rank the order in which they turn to each knowledge source when seeking information to get their job done. Rankings ranged from 1 to 9 (the 8 knowledge sources specifically identified, plus “Other”). For the purposes of analysis, a ranking of 10 was used whenever a participant did not specify a ranking for a knowledge source, under the assumption that it meant that the participant used the source less than all sources he/she bothered to rank. Figure 5.3 shows the mean ranking for each knowledge source. In this case, a lower mean value indicates that the knowledge source is used more effectively than those with a higher mean value:
The mean ranking results support the mean usage results. The Internet was ranked the highest, followed by co-workers, and then external publications. Statistical significance in the difference in mean rankings between sources would enable one to clearly differentiate the value of one source over the other. Using a one-sample T-test to test for statistical significance in the difference of mean rankings revealed that the Internet and co-workers were not significantly different, but that both had mean rankings that were significantly lower than all of the other knowledge sources ($p < .05$).

Using the co-workers knowledge source as representative of pair programming, both mean usage and the mean ranking evaluations support the hypothesis that pair programming has the potential to be an effective Knowledge Management technique; workers already consider their co-workers a prime source of information. The evidence
indicates that people turn to the Internet and co-workers significantly more often than other sources when seeking knowledge to get their job done.

5.1.2. Knowledge Foundation

Another aspect related to knowledge sources that was investigated was to examine what sources contribute the most towards building the knowledge foundation required to perform one’s job. Participants were asked to rank the following sources in the order in which they obtained the most knowledge to get their day-to-day job done:

- Academic education, such as college or high school.
- Professional education, consisting of formal classroom and online training provided by the participant’s company or organization.
- Self-education obtained through sources such as books, manuals, magazines, and the Internet.
- Mentoring or working with a partner.
- On-the-job training, in which knowledge is obtained by performing work and asking people questions when help is required.

For the purposes of this research, mentoring or working with a partner was considered to be the source most closely related to the practice of pair programming. As with the analysis of knowledge sources, the analysis of knowledge foundation sources focused on responses across all participants, regardless of the amount of pair programming performed, as a means for assessing the overall effectiveness of sources most closely related to pair programming (i.e., mentoring). Analysis of the unique ways
in which pair programmers build their knowledge foundation is beyond the scope of this research.

Figure 5.4 shows the mean ranking for each source. Rankings ranged from 1 (highest) to 6 (lowest) (the 5 sources specifically identified, plus “Other”). For the purposes of analysis, a ranking of 7 was used whenever a participant did not specify a ranking for a source, under the assumption that it meant that the participant felt that the source’s contribution was less than all sources he/she bothered to rank. It should be noted that a lower mean value indicates that the source contributes more to the participant’s knowledge foundation than those with a higher mean value:

![Knowledge Foundation Source Ranking](image-url)

**Figure 5.4. Mean Ranking of Knowledge Foundation Sources**

The results indicate that self-education contributes most towards building the knowledge foundation required to perform one’s job, followed by on-the-job training and
mentoring. The sources that contribute least to one’s knowledge foundation are professional education and “Other”. Statistical significance in the difference in mean rankings between sources would enable one to clearly differentiate the value of one source over the other. A one-sample T-test was used to determine whether or not the difference in mean ranking across different knowledge foundation sources was statistically significant. The analysis revealed that the difference in mean ranking between self-education and on-the-job training was not statistically significant, but that the difference in mean ranking between on-the-job training and mentoring was statistically significant ($p < .05$), and the difference in mean ranking between mentoring and all other sources was significant ($p < .05$).

In this case, the results do not make a strong case for pair programming as an effective Knowledge Management technique for building the knowledge foundation people use to get their job done. Mentoring, which is the knowledge foundation source that most closely resembles pair programming, was ranked number three of the five sources that were specifically identified. Knowledge gained through self-education and on-the-job training appears to contribute more to the knowledge foundation than pair programming.

5.1.3. Knowledge Management Effectiveness

The last set of questions related to pair programming and Knowledge Management effectiveness examined whether or not pair programming has a positive influence on the Knowledge Management practices of a company or organization. The analysis that has been discussed thus far has not required that the respondent’s pair
programming be taken into consideration. However, to determine whether or not pair programming has a positive influence on the Knowledge Management practices of a company or organization, the amount of pair programming performed by respondents must be factored into the analysis.

The survey included five questions intended to measure the overall effectiveness of the Knowledge Management practices of the respondent’s company or organization:

11. My company or organization is good at creating new knowledge through its people and technological resources.

12. My company or organization is good at finding, organizing, and documenting the knowledge it already possesses.

13. My company or organization is effective at acquiring knowledge from outside sources, such as consultants or products.

14. My company or organization is effective at making knowledge accessible to those who need it, when they need it.

16. How much improvement in the quality and productivity of your work could be gained by improvements to your company’s or organization’s management of knowledge?

In the research analysis, it was desirable to assess the overall effectiveness of Knowledge Management practices by aggregating the responses to these questions into an effectiveness index. A reliability analysis was performed against these five questions to determine whether or not it was statistically valid to aggregate the responses into such an index. Cronbach’s alpha test yielded a value of 0.71, which deems the questions
sufficient for aggregation. In order to aggregate them, responses to the last question, consisting of four possible responses (“A great deal,” “Some,” “A little,” and “None”), had to be converted to a format consistent with the other four questions, each of which had five possible responses (“Strongly agree,” “Agree,” “Neither agree nor disagree,” “Disagree,” and “Strongly disagree”). This was achieved by spreading the four possible responses for the last question evenly over the range of the five possible responses for the other questions, as shown in Table 5.1. Note that the order of the possible responses for the last question had to be inverted so that “A great deal,” implying there is great room for improvement, corresponded to a value reflecting weak Knowledge Management practices, and “None,” implying the organization is already at full potential, corresponded to a value reflecting strong Knowledge Management practices.

Table 5.2. Knowledge Management Effectiveness Responses and Values

<table>
<thead>
<tr>
<th>Responses for Questions 11, 12, 13, and 14</th>
<th>Responses for Question 16</th>
<th>Numerical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly agree</td>
<td>None</td>
<td>2</td>
</tr>
<tr>
<td>Agree</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>A little</td>
<td>0.666667</td>
</tr>
<tr>
<td>Neither agree nor disagree</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Some</td>
<td>-0.666667</td>
</tr>
<tr>
<td>Disagree</td>
<td></td>
<td>-1</td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>A great deal</td>
<td>-2</td>
</tr>
</tbody>
</table>
Using these values, responses from the five Knowledge Management effectiveness questions were aggregated into one “Knowledge Management Effectiveness Index”. The higher the Knowledge Management Effectiveness Index, the higher the participant rated the overall effectiveness of his/her company’s or organization’s Knowledge Management practices. The Knowledge Management Effectiveness Index for each group of respondents, grouped by the amount of pair programming that they do, is shown in Figure 5.5:

![Knowledge Management Effectiveness Index](image)

**Figure 5.5. Knowledge Management Effectiveness Index**

As can be seen, the groups had a general upward trend in the Knowledge Management Effectiveness Index as the amount of pair programming increased, with the exception of the group of respondents who pair program between 25% and 50% of the
time. To test whether there was a statistically significant positive correlation between the amount of pair programming and the Knowledge Management Effectiveness Index, Spearman’s rho was applied to the response data. The Spearman correlation indicated a positive correlation ($r_s = 0.18$), but not at a statistically significant level ($p < .10$).

Since the aggregated analysis did not yield statistically significant results, analysis was performed against each individual question to determine if there were any statistically significant correlations. Two questions yielded statistically significant results.

The analysis of question 14 revealed a statistically significant positive correlation between the amount of pair programming performed by the respondent, and the effectiveness of the knowledge dissemination in the respondent’s company or organization. Figure 5.6 shows the mean response for each group of respondents, grouped by the amount of pair programming that they do:
With the exception of the group of respondents that pair program between 25% and 50% of the time, it can be seen that there is a strong upward trend. Using Spearman’s rho against the response data revealed a positive correlation ($r_s = 0.30$) that is statistically significant ($p < .01$). The findings indicate that knowledge dissemination effectiveness increases significantly as the amount of pair programming increases.

Analysis of question 16 also yielded statistically significant results, indicating that there is less room for improvement in the Knowledge Management practices of an organization as the amount of pair programming increases. Using the response values for question 16 described in Table 5.2, the mean responses for each group of respondents, grouped by the amount of pair programming that they do, were analyzed. The results are shown in Figure 5.7. Note that because of the inversion associated with the response
values in Table 5.2, a higher mean response indicates less room for improvement, which indicates a higher degree of effectiveness:

![Knowledge Management Effectiveness (Inversion of Room for Improvement)](image)

**Figure 5.7. Inversion of Room for Improvement**

The group of respondents that pair program the most, spending between 75% and 100% of their day pair programming, responded far more positively, indicating that there is less room for improvement in their Knowledge Management practices than the other groups. A test for correlation using Spearman’s rho yielded a positive correlation ($r_s = 0.37$) that is statistically significant ($p < .01$). The results indicate that as the amount of pair programming increases, the room for improvement in the Knowledge Management practices of the organization significantly decreases.
Although the analysis of the aggregated responses did not produce statistically significant results, it did show a positive correlation between pair programming and Knowledge Management effectiveness. Further analysis of the individual questions did yield statistically significant results supporting the hypothesis that pair programming is an effective means of knowledge dissemination that has a positive influence on the Knowledge Management practices of a company or organization.

5.2. Pair Programming and Knowledge Hoarding

Two questions on the survey were designed to test the hypothesis that pair programming reduces the tendency that people have to hoard knowledge:

17. I feel the unique knowledge I possess enhances my competitive advantage over my peers when it comes to job promotions, leadership opportunities, and awards.

18. I am willing to share all of the knowledge I possess relevant to my job with my organization.

A reliability analysis was performed against these two questions to determine whether or not it was statistically valid to aggregate the responses. While both presented responses using the Likert scale previously discussed, each question approached the knowledge hoarding aspect from a different direction. A “Strongly agree” response to question 17 indicated a strong tendency to hoard knowledge, and a “Strongly agree” response to question 18 indicated a strong tendency to share knowledge. Before aggregation could be considered, the responses to one question would have to be inverted, so that numerical responses for both questions indicated the same tendency.
After inverting the responses for question 17, Cronbach’s alpha test yielded a value of -0.19, far below an acceptable value for aggregation. It is unknown whether the wording of the questions contributed to such a low alpha result, or if the questions truly measured different things.

Because aggregation was not valid, each question had to be analyzed independently of the other. The same numerical values were used for the Likert scale responses as previously discussed, ranging from –2 to 2. Testing the hypothesis required that the amount of pair programming performed by respondents be factored into the analysis of each question.

Figure 5.8 shows the mean response for each group of respondents, grouped by the amount of pair programming that they do, to question 17, regarding knowledge enhancing the respondent’s competitive advantage over his/her peers. Note that a lower mean indicates less tendency to hoard knowledge:
The results show a downward trend indicating that as the amount of pair programming increases, the tendency to hoard knowledge decreases, except for the group that pair programs between 75% and 100% of the time. Applying Spearman’s rho did indeed show an overall negative correlation ($r_s = -0.16$), but not at a statistically significant level ($p < .14$).

The same analysis procedure was applied to question 18, regarding a willingness to share knowledge. The results are shown in Figure 5.9. In this case, a higher mean indicates less tendency to hoard knowledge:
No obvious trend emerges from the resulting analysis. If pair programming reduced the tendency in people to hoard knowledge, the trend would have been upward. Applying Spearman’s rho revealed a slightly negative correlation ($r_s = -0.08$) between willingness to share knowledge and the amount of pair programming performed, but not statistically significant ($p < .47$).

Overall, the results were inconclusive regarding the hypothesis that pair programming reduces the tendency that people have to hoard knowledge. Analysis of question 17 showed weak support for the hypothesis, and analysis of question 18 showed weak opposition. Neither question yielded statistically significant results.
5.3. Pair Programming and Employee Turnover

The analysis of the survey data regarding the hypothesis that pair programming reduces the impact of employee turnover was centered around two questions:

20. If you were to leave your position tomorrow, how long would it take your organization to compensate for the loss of knowledge you possess?

21. If the most senior technical member of your organization were to leave his/her position tomorrow, how long would it take your organization to compensate for the loss of knowledge he/she possesses?

Once again, a reliability analysis was performed against these two questions to determine whether or not it was statistically valid to aggregate the responses. Cronbach’s alpha test yielded a value of 0.79, sufficient for aggregation.

Each question presented six possible responses in terms of categories of length of time. While each question presented identical responses, the duration of time for each category was not evenly distributed. Rather than using a value from 1 to 6 to represent the category that was selected, the analysis was performed using values that more accurately captured the impact of employee turnover in terms of length of time, as shown in Table 5.3:
Table 5.3. Employee Turnover Responses and Values

<table>
<thead>
<tr>
<th>Response</th>
<th>Numerical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than one week</td>
<td>0</td>
</tr>
<tr>
<td>Between one week and one month</td>
<td>0.5</td>
</tr>
<tr>
<td>Between one month and three months</td>
<td>2</td>
</tr>
<tr>
<td>Between three months and six months</td>
<td>4.5</td>
</tr>
<tr>
<td>Between six months and one year</td>
<td>9</td>
</tr>
<tr>
<td>More than one year</td>
<td>12</td>
</tr>
</tbody>
</table>

While this approach somewhat minimizes impact beyond one year, it yields more meaningful information regarding the impact of employee turnover than if a simple numerically increasing category number were used.

Applying these values, responses from the two employee turnover questions were aggregated into one “Employee Turnover Impact Index”. While not exact, the Employee Turnover Impact Index is roughly the number of months it takes for a company or organization to compensate for the loss of knowledge it incurs when an employee leaves; the higher the Employee Turnover Impact Index, the greater the impact. Thus, an indication of successful Knowledge Management is a lower Employee Turnover Impact Index.

Testing the hypothesis that pair programming reduces the impact of employee turnover required that the amount of pair programming performed by respondents be factored into the analysis of the results. Figure 5.10 shows the mean Employee Turnover
Impact Index for each group of respondents, grouped by the amount of pair programming that they do. Note that a lower Employee Turnover Impact Index indicates less impact:

![Employee Turnover Impact Index](image)

Figure 5.10. Employee Turnover Impact Index

The results show evidence of a general downward trend in the Employee Turnover Impact Index as the amount of pair programming increases. Only those groups pair programming at least 50% of the time achieved an Employee Turnover Impact Index of less than three. To test whether there was a statistically significant negative correlation between the amount of pair programming and the Employee Turnover Impact Index, Spearman’s rho was applied to the response data. The Spearman correlation indicated a negative correlation ($r_s = -0.12$), but not at a statistically significant level ($p < .26$).
Since the aggregated response did not produce statistically significant results, further analysis was performed against the individual questions. This analysis also did not yield any statistically significant results.

The analysis supports the hypothesis that pair programming reduces the impact of employee turnover, but not at a statistically significant level.

5.4. Subjective Responses

Two of the questions on the survey asked participants to comment on the strengths and weaknesses of the Knowledge Management practices of their company or organization. Here are some of the more interesting comments culled from the responses:

“We have a one hour group meeting every week in which we discuss the research going on in our group and everyone gets a chance to speak or present the work he/she is doing. This really helps in knowledge dissemination as everyone knows what other people within the group are doing.”

“Most people prefer not to contribute to the knowledge bases available. There is not enough incentive. All that matters is having running code.”

“…no one can get to my organization’s knowledge except via me. I am the repository. That is neither efficient…nor reliable.”
“...we are not yet psychic and cannot read minds, although we would be happy to dump the contents of our minds to disc and upload them to anyone who wanted them.”

“...I am amazed anything gets done.”

“It is all based on a few very brilliant and hardworking persons. If they all get hit by a bus when traveling to or from a client site, the company would need to shut down. And the funny part? They know this!”

“Knowledge is a communal property”

“We downsized before capturing knowledge and have outsourced to organizations which have little knowledge of our business. This is a huge problem.”

“...it is sometimes surprising what the others know (and don’t know)."

And finally, this comment from one of the respondents who spends between 75% and 100% of his/her day pair programming:

“Some co-workers perceive it as a disadvantage, that we don’t formally record knowledge. They feel guilty and unprofessional. I predict that soon all of us
realize that promiscuous pairing is a much more efficient approach for knowledge management.”
6. CONCLUSIONS

A survey of a cross sectional sample of programmers in information technology research and industry was conducted to examine the effectiveness of pair programming as a Knowledge Management technique. Specifically, the following hypotheses were tested:

- Pair programming reduces the tendency that people have to hoard knowledge
- Pair programming reduces the impact of employee turnover
- Pair programming is an effective means of knowledge dissemination and knowledge retention that has a positive influence on the Knowledge Management practices of a company or organization

In all, 95 responses to the survey were received and analyzed based on the amount of pair programming performed by each respondent. Results of the analysis were inconclusive regarding the hypothesis that pair programming reduces the tendency that people have to hoard knowledge, offering no statistically significant evidence to either support or contradict the hypothesis. Analysis of the survey results showed support for the hypothesis that pair programming reduces the impact of employee turnover, although not at a statistically significant level.

And finally, the survey results yielded statistically significant evidence that pair programming is an effective means of knowledge dissemination and knowledge retention that has a positive influence on the Knowledge Management practices of a company or organization. This suggests that organizations can benefit by adopting a pair programming approach to their software development processes.
Future research related to this topic should focus on identifying and examining the unique methods employed by pair programmers that have a positive affect on the Knowledge Management practices of their company or organization. Through such an evaluation, techniques could potentially be identified that could be adopted by non-pair programming organizations to improve their Knowledge Management effectiveness.
7. LIST OF REFERENCES


8. APPENDICES
8.1. Knowledge Survey

Knowledge Survey

From: David W. Palmieri
dwpalmie@unity.ncsu.edu

Introduction: I am a graduate student in Computer Science at North Carolina State University in Raleigh, North Carolina, USA. As part of my Master’s thesis research, I am gathering information about the way knowledge is acquired, disseminated, and maintained by companies and organizations in technology research, development, and service. I am interested in obtaining your input on this subject through this survey. Your participation in this survey will be anonymous, and no one will contact you for additional information or solicitation. The survey takes approximately 15 minutes to complete. Responses will be collected until Monday, February 18th, 2002.

Before filling out the survey, I encourage you to forward it to other colleagues, particularly those in other organizations, so that it reaches as wide an audience as possible.
Instructions: Most of the following questions contain possible answers from which you can choose. Please indicate your choice by placing a number or character within the square brackets (“[ ]”) that appear next to your desired choice. For questions that require you to rank the possible choices, please indicate your ranking by placing a number within the square brackets next to each choice. Additionally, some questions provide extra space (indicated by “………………..”) to allow you to specify an answer that is not provided by the available choices. There is no limit to the length of the answer you can provide.

When you have completed the survey, please return it to dwpalmie@unity.ncsu.edu. I thank you in advance for your participation.

Section 1: Knowledge Sources

When a person requires some additional knowledge or information to get their job done, they sometimes turn to various sources for the information they are seeking. When you are faced with that situation, how often do you find the information you are looking for in the following sources:

1. Internal publications (such as guidelines or “How-To” documents, not design documents) produced by your company or organization.
[ ] My company or organization does not produce internal publications of this sort.

[ ] Often.

[ ] Sometimes.

[ ] Rarely or never.

2. Design documents, from either the project at hand or a previous project, produced by your company or organization.

[ ] My company or organization does not produce design documents.

[ ] Often.

[ ] Sometimes.

[ ] Rarely or never.

3. External publications, such as books, manuals, journals, or magazines.

[ ] My company or organization does not permit the use of external publications.

[ ] Often.

[ ] Sometimes.

[ ] Rarely or never.

4. The Internet.

[ ] My company or organization does not permit the use of the Internet.
5. Co-workers within my immediate team or organization.

[ ] I do not work with others.

[ ] Often.

[ ] Sometimes.

[ ] Rarely or never.

6. Courses, either classroom or online.

[ ] My company or organization does not provide a means for allowing its employees to take courses.

[ ] Often.

[ ] Sometimes.

[ ] Rarely or never.

7. A skills “Yellow Pages” that allows me to find people within my company or organization based on skill.

[ ] My company or organization does not have a skills “Yellow Pages”.

[ ] Often.

[ ] Sometimes.
8. A database or groupware, such as SAP or Lotus Notes.
   [ ] My company or organization does not use databases or groupware to record knowledge or skills-related information.
   [ ] Often.
   [ ] Sometimes.
   [ ] Rarely or never.

9. Please rank the following in the order you are MOST LIKELY to turn when seeking additional information to get your job done. Place a 1 next to the source you are most likely to go to first, a 2 next to the source you are most likely to go to second, and so on.
   [ ] Internal publications (such as guidelines or “How-To” documents, not design documents) produced by your company or organization.
   [ ] Design documents, from either the project at hand or a previous project, produced by your company or organization.
   [ ] External publications, such as books, manuals, journals, or magazines.
   [ ] The Internet.
   [ ] Co-workers within my immediate team or organization.
   [ ] Courses, either classroom or online.
[ ] A skills “Yellow Pages” that allows me to find people within my company or organization based on skill.

[ ] A database or groupware, such as SAP or Lotus Notes.

[ ] Other (Please specify): ………………..

[ ] Other (Please specify): ………………..

[ ] Other (Please specify): ………………..

10. Please rank the following in the order in which you obtained the most knowledge to get your current, day-to-day job done.

[ ] Academic education (e.g., college or high school).

[ ] Professional education (i.e., formal classroom or online training provided by your company or organization).

[ ] Self-education (e.g., books, manuals, magazines, the Internet, etc.).

[ ] Mentoring or working with a partner.

[ ] On-the-job training in which you learn the job as you do it, asking people questions when you need help.

[ ] Other (Please specify): ………………..

[ ] Other (Please specify): ………………..

[ ] Other (Please specify): ………………..

Section 2: Knowledge Acquisition, Dissemination, and Maintenance
11. My company or organization is good at creating new knowledge through its people and technological resources.

[ ] Strongly agree.
[ ] Agree.
[ ] Neither agree nor disagree.
[ ] Disagree.
[ ] Strongly disagree.

12. My company or organization is good at finding, organizing, and documenting the knowledge it already possesses.

[ ] Strongly agree.
[ ] Agree.
[ ] Neither agree nor disagree.
[ ] Disagree.
[ ] Strongly disagree.

13. My company or organization is effective at acquiring knowledge from outside sources, such as consultants or products.

[ ] Strongly agree.
[ ] Agree.
[ ] Neither agree nor disagree.
[ ] Disagree.
14. My company or organization is effective at making knowledge accessible to those who need it, when they need it.
   [ ] Strongly agree.
   [ ] Agree.
   [ ] Neither agree nor disagree.
   [ ] Disagree.
   [ ] Strongly disagree.

15. Does your company or organization have personnel in place specifically responsible for managing knowledge (e.g., Chief Knowledge Officer, Knowledge Project Manager, Knowledge Management Specialist, Knowledge Team, etc.)?
   [ ] Yes.
   [ ] No, or not that I am aware of.

16. How much improvement in the quality and productivity of your work could be gained by improvements to your company’s or organization’s management of knowledge?
   [ ] A great deal.
   [ ] Some.
   [ ] A little.
17. I feel the unique knowledge I possess enhances my competitive advantage over my peers when it comes to job promotions, leadership opportunities, and awards.

[ ] Strongly agree.
[ ] Agree.
[ ] Neither agree nor disagree.
[ ] Disagree.
[ ] Strongly disagree.

18. I am willing to share all of the knowledge I possess relevant to my job with my company or organization.

[ ] Strongly agree.
[ ] Agree.
[ ] Neither agree nor disagree.
[ ] Disagree.
[ ] Strongly disagree.

19. In your company or organization, how much are rewards based on individual technical accomplishments, versus team success?

[ ] Virtually all individual accomplishments.
[ ] Primarily individual accomplishments, but also some team success.
Equally individual accomplishments and team success.

Primarily team success, but also some individual accomplishments.

Virtually all team success.

20. If you were to leave your position tomorrow, how long would it take your organization to compensate for the loss of knowledge you possess? (Please make your best guess)

Less than one week.

Between one week and one month.

Between one month and three months.

Between three months and six months.

Between six months and one year.

More than one year.

21. If the most senior technical member of your organization were to leave his/her position tomorrow, how long would it take your organization to compensate for the loss of knowledge he/she possesses? (Please make your best guess)

Less than one week.

Between one week and one month.

Between one month and three months.

Between three months and six months.

Between six months and one year.
Section 3: General

22. How long have you been a programmer in industry or research?
   [ ] I am not a programmer in industry or research.
   [ ] Less than one year.
   [ ] Between one year and two years.
   [ ] Between two years and five years.
   [ ] Between five years and ten years.
   [ ] More than ten years.

23. How long have you been with your current employer?
   [ ] Less than one year.
   [ ] Between one year and two years.
   [ ] Between two years and five years.
   [ ] Between five years and ten years.
   [ ] More than ten years.

24. How satisfied are you with your job?
   [ ] Very satisfied.
   [ ] Somewhat satisfied.
Neither satisfied nor dissatisfied.

Somewhat dissatisfied.

Very dissatisfied.

25. What software methodologies do you use in your current organization? Please indicate all that apply.

I do not develop software.

Waterfall Model.

Spiral Model.

Rapid Application Development (RAD).

Computer Aided Software Engineering (CASE).

Extreme Programming (XP).

None / ad hoc.

Other (Please specify): ………………..

26. Approximately how much of your day is spent collaborating with others?

Less than 10%.

Between 10% and 25%.

Between 25% and 50%.

Between 50% and 75%.

More than 75%.
27. Pair programming is a practice in which two programmers work side-by-side at one computer, collaborating on the same design, algorithm, or test.

Approximately how much of your day is spent pair programming?

[ ] Less than 10%.
[ ] Between 10% and 25%.
[ ] Between 25% and 50%.
[ ] Between 50% and 75%.
[ ] More than 75%.

28. (Optional) What do you consider to be the strengths of your company’s or organization’s management of knowledge?

…………………………………………………………………………..
…………………………………………………………………………..
…………………………………………………………………………..

29. (Optional) What do you consider to be the weaknesses of your company’s or organization’s management of knowledge?

…………………………………………………………………………..
…………………………………………………………………………..
…………………………………………………………………………..

…………………………………………………………………………..
Thank you for participating in the survey. Please return the completed survey to dwpalmie@unity.ncsu.edu.