Big data and ICTs for human-capabilities:

Opportunities and challenges for skills- and human-development through the use of information & communication technologies (ICTs) and data-intensive science (big data) in TVET and the world of work

Human Development

Information Technologies

Communication Technologies

World of Work & TVET

Psychology

Skills

Working paper

www.fwdi.org

Alexander Gloss
North Carolina State University

Tara S. Behrend
George Washington University

David Blustein
Boston College

In collaboration with

Lori L. Foster
North Carolina State University

Borhene Chakroun
United Nations Educational, Scientific, and Cultural Organization

Special thanks:

Special thanks to Mr. Robert Daniel for his work alongside Mr. Borhene Chakroun in the development of the concept note preceding this working paper. Special thanks also to Professors Stuart C. Carr and Dharm P. Bhawuk who worked with the authors on the United Nations Development Programme's (2014) report "The role of the private sector in inclusive development: Barriers and opportunities at the base of the pyramid" and whose insight helped contribute to ideas around the positive skills cycle.

Abstract

Executive Summary

1. Information and communication technologies (ICTs) have allowed for and accelerated the development of big forms of digital data in the knowledge economy

The 21st century economy has been increasingly shaped by the importance of knowledge – this includes the importance of generating and using information and the importance of communication and collaboration on local and global levels. The knowledge economy has been facilitated, and encouraged, by the widespread prevalence of information (e.g., cloud computing) and communication technologies (e.g., mobile phones). Notably, the prevalence of ICTs has extended into lower-income settings. Today, the vast majority of people in lower-income settings possess some degree of access to mobile-phone technologies. Along with the rising prominence of ICTs, the world has seen a dramatic increase in the "intensity" of digitized data, that is, an increase in its amount, an increase in the rate at which it is created, and/or an increase in the complexity in, or lack of clarity to, its structure. The intensity of digital data has posed many opportunities but has also led to difficulties in understanding its validity and reliability and difficulties in effectively harnessing it effectively for insight and practical change. Advanced technological and scientific methods to acquire, manage, and analyze big data have been under development in a wide range of disciplines for decades and we refer to them broadly as forms of "data-intensive science."

2. Human capabilities are impeded by barriers to skills development in technical and vocational education and training (TVET) and in the workplace

Success in the 21st century knowledge economy relies not only on the creation or possession of knowledge, but on the implementation of that knowledge in the world of work. The skills necessary to generate knowledge and to convert knowledge into work outcomes have often been labeled "21st century skills" due in part to their relevance to the increasingly digital world of work. Yet skills are not only relevant to work. As illustrated by the importance of lifelong learning, skills are important to broader human development because they form major components of human capabilities, that is, components of the freedom of people to pursue their goals in life. Continued progress in promoting human welfare and economic prosperity in lower-income settings, with marginalized populations, and in the informal economy relies in large part on overcoming barriers to skills development in technical and vocational education and training (TVET) and in the workplace. These barriers include social, political, economic, and psychological forces on global, national, workplace/institutional, and individual levels.

3. Economic, social, and political barriers to skills development also have psychological implications

Because access to information, and power, is increasingly a question of whether, and how well, people interface with ICTs, many barriers and enablers to economic success come from people's relationships with technology, education, training, and the world of work. Despite the

tremendous potential for combinations of data-intensive science and ICTs to promote skills- and human- development, relatively little attention has been focused on their potential. Arguably even less attention has been placed on the unique potential of psychology to help ensure that data-intensive science and ICTs help to promote enhanced human capabilities for populations that are often left behind in the knowledge economy.

4. This working paper presents an agenda for skills- and human development: Information and communication technologies for human capabilities (ICT4H)

In light of the above observations, we propose that combinations of data-intensive science, ICTs, and methods and theories from the psychological sciences are uniquely suited to promoting skills and broader human-capabilities in lower-income settings, in the informal economy, and for marginalized populations. We refer to the combination of data-intensive science with ICTs and psychological insights and methods for the purpose of developing work skills and human capabilities as “information and communication technologies for human-capabilities” (ICT4H).

In this working paper, we: 1. provide a broad conceptual overview and operational definitions for “big data,” human development, ICTs, and skills (Section 1), 2. Review the overlap of these key concepts with one another (Sections 2-4), 3. Describe the unique characteristics of and potential for ICT4H (Section 5), and 4. Present salient challenges and considerations for ICT4H along with recommendations for a future ICT4H agenda (Section 6).

Table of Contents

1. Concepts and Definitions

Below, we provide operational definitions for the terms “big data,” “skills,” “information and commination technologies,” and “development” for use in this working paper. These terms carry with them a variety of connotations and denotations from a wide number of academic disciplines and professional sectors. In addition, the recent prominence of the concept of big data has led to confusion regarding its typical characteristics and potential.

1.1 Big data

1.1.1 What is “big” data?

In line with definitions and treatments from multiple sectors (see Chen, Mao, Zhang, & Leung, 2014; IBM, 2014; McAfee & Brynjolfsson, 2012; McKinsey & Company, 2011; World Bank, 2014a), the term “big data” can be interpreted as digital data that exhibit one or more qualities that necessitate “non-traditional” data acquisition, management, and/or analysis. These qualities can be defined as the data’s:

- Volume: the size of data in terms of bytes;
- Velocity: the speed at which data is created and/or loses currency;
- Variety: data’s structural complexity, its lack of structure, and the connections of one or more data sub-sets to other sub-sets; and
- Veracity: the reliability of the data and/or its validity for an intended use.

Throughout this working paper, we refer to data that exhibits one or more of these qualities as “big data.”

1.1.2 Data-intensive science

The use of big data in a rigorous manner requires a specific form of scientific investigation that we refer to as data-intensive science. Data-intensive scientific investigation involves at least four differentiable sub-steps:

- The generation of big data;
- Big-data acquisition;
- Big-data management;
- Big-data analysis.

We briefly review major components of each of these four sub-steps in the sections that follow.

1.1.3 Generating big data

As mentioned above, big data can vary according to its volume, velocity, variety, and veracity. The creation of such data can emerge due to an intervention (e.g., through the creation of an online survey or the wearing of a digital device that collects biometric information) or through the normal operation of a system (e.g., the operation of a mobile phone network). Beyond this distinction, it is helpful to point out that the generation of big data often occurs from one or more of the following:

- The interface between a physical environment and technological devices (e.g., temperature sensors);
- Digital connections between technological devices (e.g., between motor vehicles and mobile phones);
- The interface between people and technological devices (e.g., mobile phones);
- Interactions among individuals in a digital environment (e.g., between people in a social network).

1.1.4 Big data acquisition

Engagement with big data often begins with attempts to collect raw data from the location where that data was generated. Depending on data scientists' ability to interface with the digital environment where the raw data is generated, different strategies can be employed for data retrieval. Two examples help to illustrate this point: when data is retrieved from digital systems that investigators have some degree of control over, they can often efficiently isolate data in a form that is relatively easy to manage and analyze. For example, those with access to a digital system like a mobile phone network can introduce log files which are pre-programmed to identify certain events in a digital system (for example, the keystrokes on a phone). If investigators do not have access to and/or knowledge of a system/application's data-extraction interface/protocols, or if data needs to be retrieved across multiple systems (for example, from multiple websites), other methods can be employed including "data scraping." Examples of data scraping including the use of web crawlers that are programmed to retrieve certain types of information from webpages and screen scraping which records the visual output of a system (for example, visual snapshots of a computer's screen).

1.1.5 Big data management

The distinction between big-data acquisition and management is often muddied as aspects of data management can take place throughout the process of acquiring raw data and then storing it in a place and manner that supports later analysis. In addition, the characteristics of data management can vary widely depending on the intended use of the data. Management designed to support the use of data by academic researchers is often far simpler than forms of data management undertaken by practitioners working to support the sustained and automated functioning of a digital system (for example, a data specialist working to support the day-to-day functioning of an e-commerce website). In the latter case, data management must often be carried out in a way that is designed to support the sustained analysis of data and the automated reintegration and use of that information in a digital system. In this working paper, we focus mostly on forms of data management that are necessary to support the use of data by academic researchers and policymakers. Typically, such stakeholders are not looking to immediately and automatically reintegrate analyzed data into a digital system for the purposes of management; however, exceptions do exist. The use of big data to support real-time indicators of social phenomena or the building of computer-adaptive online module are just two examples of potential uses of big data that would necessitate a more complex approach to data management.

When big data is acquired, it must be managed in a series of iterative, overlapping, and repeated sub-steps including integration, cleaning, and storage. Integration frequently involves the processes of extraction, transformation, and loading (ETL) which include the selection of data, the application of rules to convert the structure of that data into a desired format, and then the loading of that data into an intermediary or final location. The cleaning of data involves the automated rule-based categorization and resolution of errors, missing data, and data redundancy. Finally, storage often involves placing big data into either a single system with sufficient processing and storage capacity or a set of distributed systems for analysis. Distributed systems utilize software infrastructures (e.g., Hadoop) designed to divide big data into blocks that are clustered across different computing systems and processed by those systems in parallel with one another.

1.1.6 Big data analysis

As with the distinction between big-data acquisition and management, the distinction between big data management and analysis is often blurred. Big-data analysis, sometimes referred to as data mining, can take place across a distributed system through parallel processing or with smaller sub-sets of data that can be accessed from across the entire system, brought into an analytical environment, and then processed by a single computer. Big data analysis generates information that can either be quickly reintegrated into a system or directed outside of a system for human consumption. In the former case, data scientists rely upon customized algorithms that automatically process and react to incoming data according to pre-assigned criteria – a process sometimes known as machine learning. In the latter case where information is transmitted to a human, learning takes through customized statistical analyses, simulations, and/or modeling. A wide diversity of big data analytics can be carried out by both specialized and more traditional analytical software frameworks (e.g., R).

As put forward by Chan et al. (2014), the use of big data for commercial applications, (e.g., e-commerce) electronic network applications (e.g., online search engines), and scientific applications (e.g., environmental research) have generated the development of major sub-fields of big data analysis, including:

- Structured data analysis;
- Text-data analysis;
- Web-data analysis;
- Multimedia-data analysis;
- Network-data analysis; and
- Mobile-traffic analysis.

Within each of these sub-fields, data-scientists rely upon familiarity with different types of data (e.g., visual or semantic), data structures, and analytical techniques. For example, within different analytical sub-fields, data scientists might utilize: social network analysis, logistic regression, Bayesian estimation methods, non-linear and/or multilevel models. However, underlying this diversity are certain common analytical goals, including:

- Describing surface-level (manifest) or underlying (latent) characteristics of data;
- Classification into categories;
- Explaining the association or relationship between two or more variables;
- Predicting events, trends, or membership in certain classes; and
- Testing models or hypotheses.

Moreover, many analytical approaches rely upon a common set of statistics to accomplish certain sub-goals, including correlational or regression-based analyses and cluster or factor analyses.

1.1.7 The semantic web

Throughout various stages of data-intensive science the structure and content of data is often mapped, categorized, and standardized in various manners. These processes are major prerequisites to, and features of, drawing meaning and usefulness from big data. Mapping, categorization, and standardization often involve the use of data formats, protocols, and meta-data models. The development of Semantic Web technologies (www.w3.org) which include movement toward a set of interrelated frameworks for data is especially relevant to issues of skills as they have helped to facilitate standardized skill vocabularies, ontologies, and the automated linking of data across databases including for the classification of European Skills, Competences, Qualifications, and Occupations (ESCO; www.ec.europa.eu). These technologies include the Resource Description Framework (RDF)--a general-purpose languages for representing information, the Simple Knowledge Organization System (SKOS) data model which supports standardization of and linking between taxonomies, and the formal knowledge Web Ontology Language (OWL).

1.1.8 A paradigm shift?

To some, data-intensive science operates according to a different paradigm than “traditional” forms of empirical observation, experimentation, and/or model/hypothesis testing. As argued by some proponents, data-intensive science utilizes new approaches made possible by greater computing power and network speeds which combine observation, theory building, experimentation, and simulation in a close, rapid, and iterative fashion (Hey, Tansley, & Tolle, 2009).

In contrast, critics of some approaches to data-intensive science accuse its practitioners of sometimes relying too heavily on observation and induction in a way that can lead to improper conclusions, misleading inferences of causality, and the inability to extrapolate beyond existing situations and samples (The Aspen Institute, 2010). The advantages and dangers of various forms of data-intensive science are discussed in greater depth later in this working paper. Regardless as to whether data-intensive research can or should rest on a new research paradigm, its practice still seems to hold the same broad goals as traditional research, namely to transform data into information that can promote understanding and inform decisions.

While the social sciences have wavered back and forth on the balance between theory and empiricism somewhat reliably for 100 years, the emergence of data-intensive science might most clearly be viewed as the latest stage in that balancing act.

1.2 Definition: Human development

A definition of human development is derived from the tradition of the human capabilities approach in international development and development economics. The human capabilities approach to development attempts to provide a universal but also culturally-sensitive conceptualization of human progress by defining wellbeing as a state where people have the actual capabilities, or freedom, to live lives they have reason to value (Sen, 1999). In other words, the human capabilities approach envisions human development as a process of expanding individuals' freedoms to meaningfully pursue their own goals as long as those goals do not undermine the freedom of others. As reflected in the United Nations Development Programme's Human Development Index (UNDP, 2014), the human capabilities approach assumes that people's freedoms are usually facilitated by establishing basic conditions in society such as a low incidence of sickness, fundamental levels of education, and a decent quality of life supported by a threshold level of income.

1.3 Definition: Information and communication technologies

According to UNESCO, the term information and communication technologies (herein ICTs) refers to technologies that are used to "transmit, process, store, create, display, share, or exchange information by electronic means" (UNESCO, 2007, p.1). As illustrated by this definition, ICTs include communication technologies (CT) like mobile phones whose primary purpose is to share and exchange information or to facilitate collaboration. In addition, ICTs also include information technologies like statistical and visualization software that acquire, manage, analyze, and present information. Throughout this working paper, we differentiate information technologies (IT) from communication technologies (CT); however, such a distinction is somewhat artificial as nearly all ICTs include both strong informational and communicative components.

1.4 Definition: Skills

A universally accepted definition of "skills" is difficult to isolate given the relevance of the term to a diversity of fields from education and development economics to psychology. In alignment with an inclusive definition of the Organization for Economic Co-Operation and Development (OECD; 2012, p.12) we define skills as individual learned capacities to perform an activity or task. First and foremost, this definition differs from other understandings of skills by establishing them as necessarily human and psychological constructs. Adopting such a definition deliberately excludes a number of concepts that are sometimes used as proxies and/or confused with skills (e.g., years of education). In addition, the requirement that a skill leads to the successful performance of an activity or task often distinguishes skills from purely declarative forms of knowledge (e.g., memorization of facts and figures). Furthermore, the requirement that skills be "learned" provides a useful distinction between skills and other characteristics that are either innate, or cannot be effectively or appropriate "taught" – for example innate physical abilities, relatively stable characteristics like personality, values, and quickly changing attitudes and affective states. Decades of research in vocational and industrial-organizational (I-O) psychology has shown that the above distinctions are important when understanding, predicting, and attempting to alter people's behavior and physical and psychosocial well-being at work (e.g., Chernyshenko & Drasgow, 2011). Vocational psychology and I-O psychology are related applied social sciences that involve the empirical study of career development, worker wellbeing, and organizational effectiveness. Finally, it is important to note that the requirement that skills be tied to tasks and activities brings that concept close to the world of work and to technical and vocational education and training (TVET).

2. The role of skills in human development

As highlighted by the United Nations Educational, Scientific, and Cultural Organization (UNESCO), the development of work-related skills is a fundamental component of global and sustainable human development (UNESCO, 2012). In this section, we review key concepts and developments emerging from a variety of disciplines and sectors relating to skills, psychology, TVET, and the world of work – both inside and outside of the context of human development. These disciplines include economics, education, management, and psychology – including community, health, vocational, and I-O psychology sub-disciplines.

While work-related skills are necessary for individuals, organizations, communities, and nations to prosper financially and economically, developing work-related skills is also a major part of lifelong learning and an important part of people's broader capabilities and freedom to pursue their goals in life. Viewed from another standpoint, and as highlighted by the psychology of working perspective, success in one's working life provides a means for survival and power, a means of social connection, and a means for psychological self-determination which is a fundamental human need (Blustein, 2006). Moreover, work-related skills are both engines to successful performance in a work environment and necessary for individuals to help navigate the social, economic, and political systems that can control access to vocational development and socioeconomic mobility (Blustein, McWhirter, & Perry, 2005). Indeed, while jobs are in many ways the cornerstone of economic and social development, the manner in which jobs impact upon personal wellbeing and broader society is largely contingent upon the skills that those jobs require, engender, and utilize (OECD, 2012; World Bank, 2012a). For example, as highlighted by a recent multinational study by the OECD (2013):

- Certain skills, like literacy, are tied to important workplace outcomes like productivity and wages;
- Differences in the use of key skills at work accounts for approximately half of the gender gap in wages; and
- Skill proficiency is closely associated with good health, participation in political processes, and levels of trust in others.

The skills necessary for success in the world of work are largely developed through four main pathways: (1) foundational education; (2) pre-employment technical and vocational education and training (TVET), (3) forms of education and training in the workplace (including informal forms gained through work experience), and (4) lifelong learning outside of education and work (OECD, 2012; UNESCO, 2012).

2.1 Barriers to skills development

As highlighted by recent reports from the United Nations Development Programme (UNDP; 2014) and the World Bank (2012b), major barriers to skills-development in vocational education and training and in the workplace are linked to socioeconomic factors in lower-income settings and within the informal economy. As supported by scholarship from vocational psychology and I-O psychology, many of these barriers might emerge from societal sources of oppression and marginalization that can undermine individual success in education, training, and the workplace

(UNDP, 2014). In particular, socioeconomic forces can help strengthen or undermine several individual attributes that support people's successful engagement with TVET and the world of work. Prominently, these attributes include:

- Sources of resilience and resistance (e.g., self-control);
- Internalized social identities (e.g., social class);
- Beliefs about oneself in an academic and work context (e.g., self-efficacy); and
- Motivational factors (e.g., expectancy of outcomes).

Positive sources of resilience, identity, self-beliefs, and motivations can enhance individual agency and can help people to overcome structural disadvantages. Moreover, positive versions of these attributes can help to support the development of skills in vocational education and training. In turn, successful skills development can help to positively affect resilience, social identities, self-beliefs, and motivations both directly (e.g., from satisfaction and self-actualization resulting from mastering skills) and indirectly through active employment or increased income and social standing.

Despite these individual attributes' potential to help people overcome obstacles in society, they are often influenced and overborn by systems and forces outside of individual control – principally often including access to social, economic, and political resources. Moreover, in the context of the world of work, major influences on individual attributes come from the workplace – either in a formal organization (e.g., a factory) or in an informal work setting (e.g., the neighborhood of a micro entrepreneur). Dynamics in the workplace are also important to consider because, as mentioned above, the workplace is a major source of skills development. Skills development in the workplace often occurs through three main avenues:

(i) The day-to-day practice of and attempted improvement on existing work skills (e.g., experimenting with new and more efficient ways of accomplishing a task);

(ii) Informal forms of workplace training and development (e.g., shadowing a more experienced employee); and

(iii) Formal forms of training and development (e.g., a structured on-the-job training program or a formal mentorship program).

A principal mechanism for each one of these forms of skills development is goal setting. In deliberate and/or unnoticed ways, people continually set goals for themselves at work, in education, and throughout life. One of the most powerful and ever-present motivators is when people set challenging yet specific goals for themselves or when others set challenging yet specific goals that those people identify with and internalize as their own (Locke & Latham, 2013). Under the right conditions, the accomplishment of challenging and specific goals at work can bring satisfaction, a greater belief in oneself, and an enhanced sense of self-determination, meaning, and impact in relation to work. Moreover, as explained above, meeting goals – either through day-to-day work activities, informal training, and formal training can produce skills development. With success meeting goals and greater skills, a virtuous "positive" work cycle can result which both develops and feeds upon people's psychological empowerment at work. Yet, the correct conditions are key and forces in the workplace and in broader society can undermine and break this positive work cycle. Major workplace influences on skills development, goal setting, and psychological empowerment include (see Seibert, Wang, & Coutright, 2011):

- Formal and informal leadership by others;
- The design and characteristics of the work to be completed; and
- Access to and utilization of human-resources support like training programs.

In addition, sources of broader social, economic, and political support can impact upon factors in the workplace and directly upon important individual attributes (e.g., access to resources). Altogether, the interaction of individual attributes, workplace factors, and broader societal support helps to shape skills development in the workplace. A simplified representation of the interrelationship between some of the major factors that affect the psychological underpinnings of skills development in the workplace and in TVET is represented in Figure 1.

Continued research into and efforts to promote skills development within lower-income settings and with marginalized populations has been called for; in particular, a better understanding of the best-practices and unique dynamics with low-skilled individuals, the long-term unemployed, marginalized social groups, and with the implementation of both TVET and workplace training in lower-income/informal environments is needed (United States Department of Labor, 2014; World Bank, 2012b). Many of these populations and settings are affected by the prevalence of informal economies that make the implementation of standard vocational, organizational, and workforce development strategies often at best questionable and sometimes inapplicable, inappropriate, or harmful (UNDP, 2014). According to the ILO (2013), forms of employment in the informal economy include jobs not subject to legal protection and regulation, social protection, or entitling job holders to critical employment conditions and benefits like decent working hours, safe working conditions, and paid-leave for sickness. Workers operating in informal and vulnerable employment situations include own-account workers, micro-entrepreneurs, contributing family workers, and members of informal cooperatives. Despite the critical importance of supporting the working welfare, skills, and empowerment of those operating in the informal economy, as detailed by the ILO (2013), “for many countries a detailed statistical knowledge of the informal economy remains at best fragmented, cursory and anecdotal.” A key weakness in the international community’s understanding of skills-development in lower-income settings, the informal economy, and with marginalized populations is the lack of knowledge around the unique processes and considerations in relation to skills development in the workplace (World Bank, 2012b). An appreciation of psychological dynamics in education and the workplace will be essential to efforts to understand the processes of skills development as on individual, workplace, and institutional levels (e.g., within TVET institutions), these intrapersonal and interpersonal processes are largely psychological in nature. However, a focus on psychological dynamics should not remove the focus on the role of broader social, economic, and political systems in limiting skills development. Instead, a focus on psychology should be used to better understand the effect of these systems on psychological wellbeing, the psychological process of skills development, and the psychological aspects of human capabilities.

Positive work and skill-development cycle

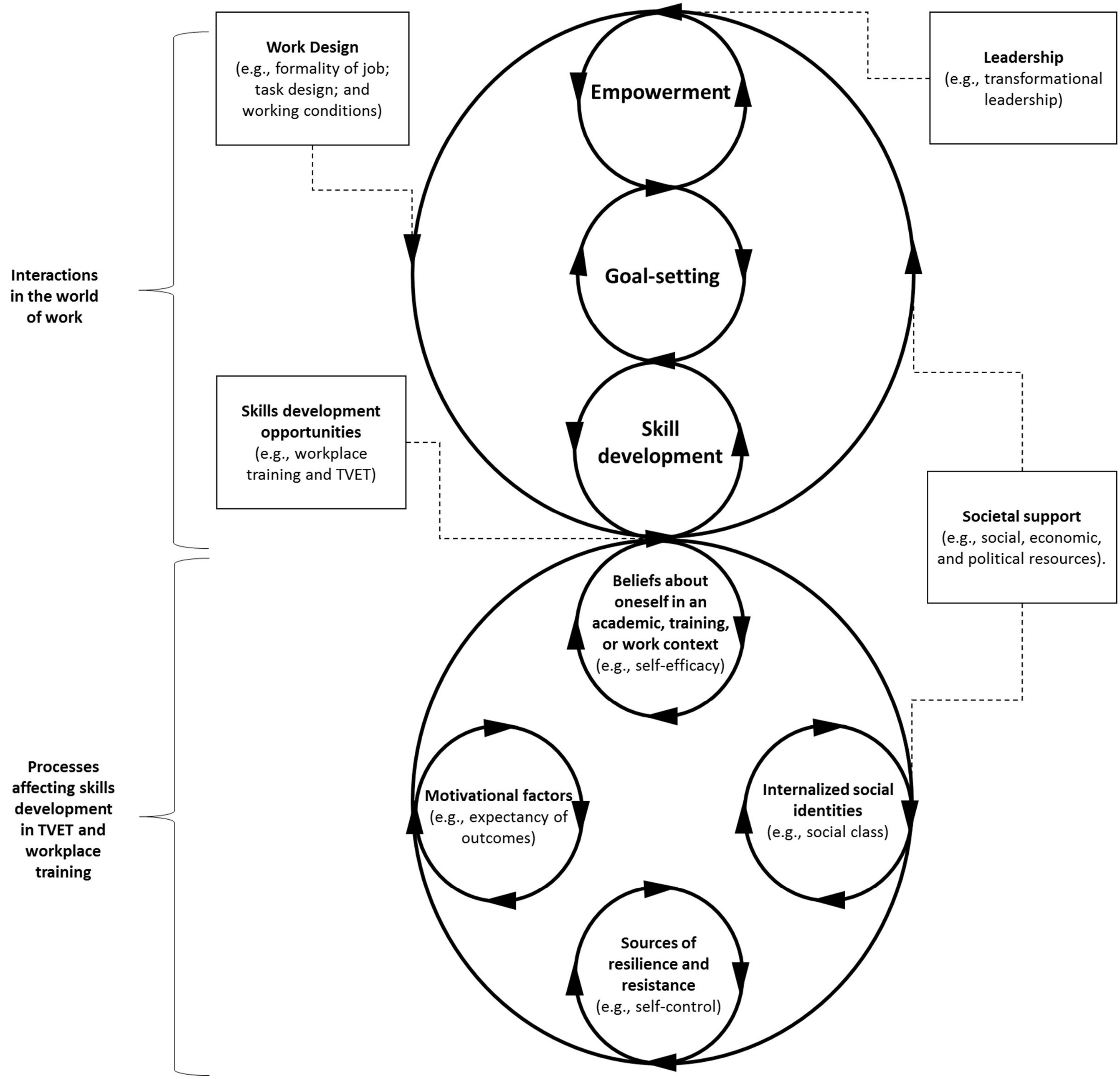


Figure 1: The positive work and skill-development cycle: factors affecting success in TVET, and other influences on the psychological underpinnings of skills development from the workplace and broader society. This figure represents a synthesis and simplification of multiple theories/frameworks regarding the psychology of working, workplace motivation, empowerment, and skill development; the placement of arrows is not meant to indicate any specific order of operations. The figure is synthesized from theories, frameworks, and/or empirical findings from Blustein 2006; Gagne & Deci, 2014; Locke & Latham, 2013; Meyer, Becker, & Vandenberghe, 2004; Salas & Canon-Bowers, 2001; Seibert et al., 2011; Spreitzer, 2008, and UNDP, 2014.

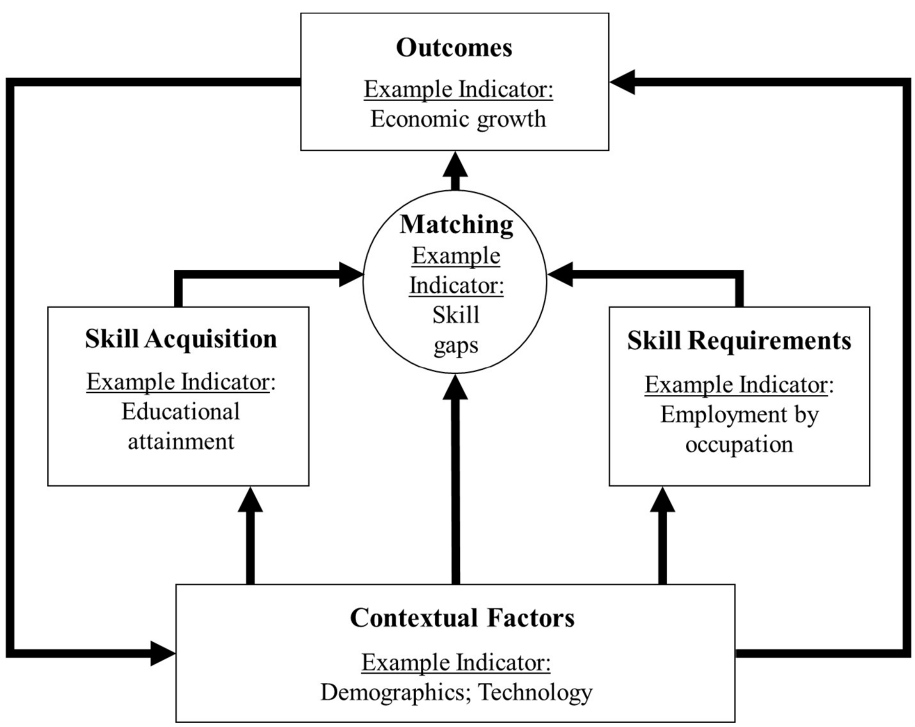
2.2 Frameworks for understanding the role of skills in human development

Significant recent attention from prominent intergovernmental organizations has been focused on the role of skills in global development. In particular, an initiative spurred by human resource development pillar of the G20 Multi-Year Action Plan has coordinated the efforts of the ILO, the OECD, UNESCO, and the World Bank to create internationally comparable skills indicators in developing countries; these indicators are meant to:

1. Better match training to existing and emerging skills needs;
2. Identify gaps in the education system;
3. Identify links between education, health, gender gaps, and life-long skills development; and
4. Produce a comparable database across countries to serve as a monitoring tool for assessing employable skills development in low-income countries.

In response to the call from the G20, the international institutions listed above have developed an online platform that serves as an international repository for information on skills (www.skillsforemployment.org) and in cooperation with the European Training Foundation these institutions have developed a conceptual framework for skills indicators (OECD, World Bank, ETF, ILO, & UNESCO, 2013). This framework (Figure 2) conceptualizes the role of skills in society and provides examples of potential indicators for major aspects of this relationship, including contextual factors affecting skill acquisition and skill requirements, the matching of skill acquisition to skill requirements, and the outcomes of that matching.

Figure 2: Conceptual framework for the role of skills in society – figure adapted from OECD et al., 2013



As more fully discussed later in this working paper, combinations of data-intensive research, communication technologies, and psychological measurement and assessment, have the potential to more exactly measure skill requirements, skill proficiency levels, and skill matching. In order to support this greater measurement potential, a common international framework that accounts for individual- and workplace-level skill dynamics and major indicators of those dynamics is needed. As an example of the rough outlines of such a framework, Gloss, Thompson, & McCallum (2015) integrated aspects of the positive work and skill-development cycle into the framework from OECD et al. (2013). As can be seen in Gloss et al.'s (2015) framework (Figure 3) the figure replaces the depersonalized "skill matching" concept with a set of interrelated psychological dynamics that drive and/or inhibit an individual's skill development.

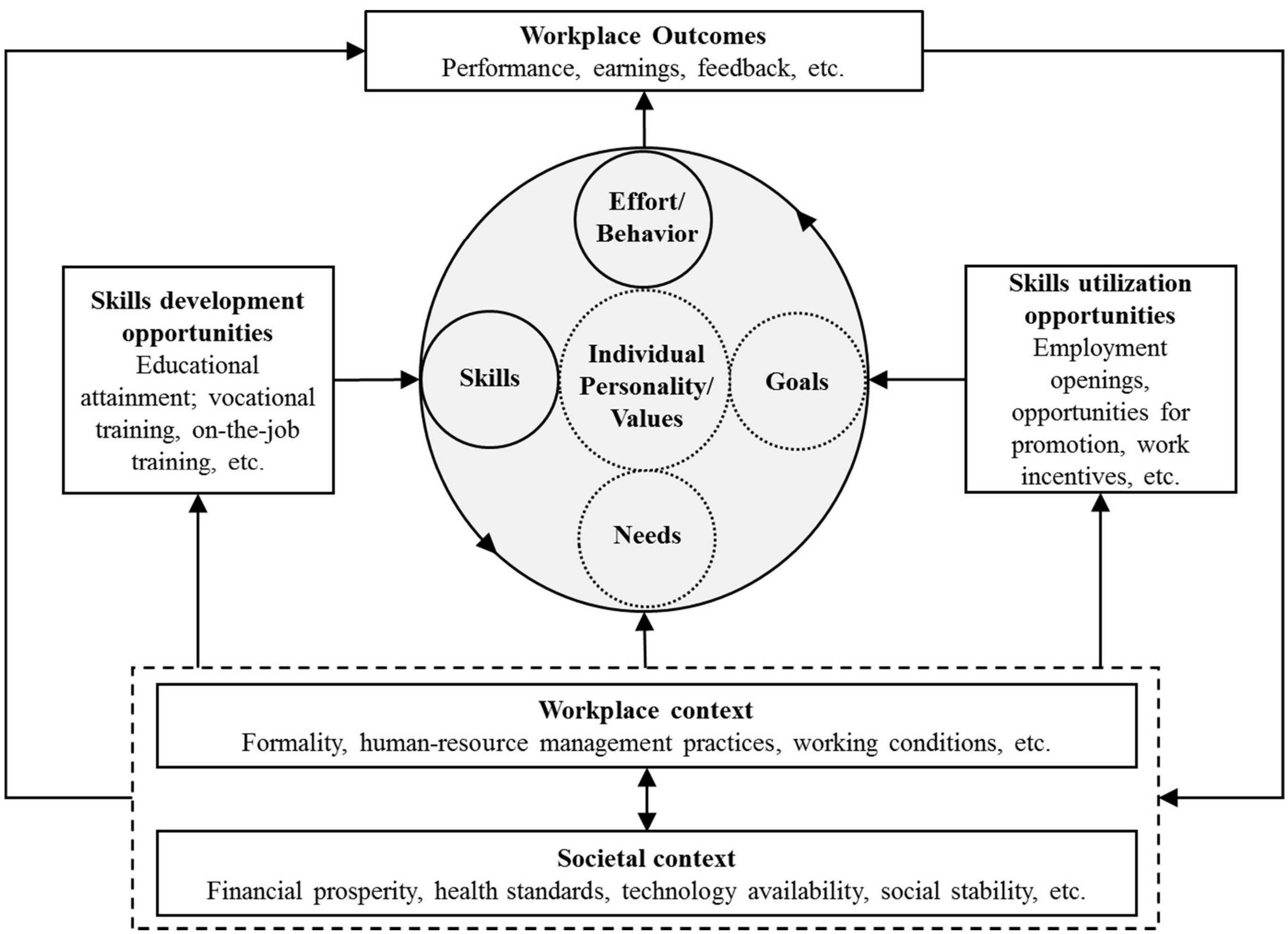


Figure 3. Framework for skills development from Gloss et al. (2015) including dynamics on the individual, workplace, and societal levels of analysis. Gray circles represent constructs/dynamics on an entirely individual-level of analysis. Figure based on a combination of theoretical frameworks from the OECD et al., (2013); UNDP (2014); Locke and Latham (2013); and Meyer, Becker, and Vandenberghe (2004).

The substitution of skill matching for an individual reflects the fact that skills-development and matching involve people with unique goals, needs, and individual personalities/values. Moreover, the process of skills development is not a simple question of matching skill needs to skill supply – instead, it is a process that involves decisions that arise from individuals' motivations and efforts to take advantage of skill-development opportunities and opportunities to utilize those skills in a variety of settings (e.g., the workplace and in TVET). Figure 6 differs from the framework provided by OECD et al. (2013) in that it attempts to incorporate major phenomena relating to skills on individual and workplace/institutional levels in addition to just a societal level of analysis. Future frameworks should more fully account for psychological dynamics in education and training as reflected in Blustein (2006)'s perspective, which emphasized the critical importance of equal access to the antecedents of a dignified work life—relevant and rigorous education, lifelong learning ,and retraining to maintain currently in the job market .

2.3 Skill taxonomies and 21st century skills

Beyond the need to establish common agreement regarding frameworks by which to measure skills, there is a need to understand useful taxonomies of skills and to concentrate attention on skills that are important in a 21st century economy. A number of ongoing international initiatives to develop such frameworks are being undertaken by national governments, international institutions, and civil-society coalitions. Prominent examples of skill-related taxonomies include the European Commission's ESCO framework which divides skills/competencies into job-specific and "transversal" categories. Within these top-level divisions, there are 25 job-specific sub-categories that relate to different knowledge and service domains (e.g., computing, law, and physical sciences) and five transversal sub-categories, namely:

- Application of knowledge;
- Attitudes and values at work;
- Language and communication;
- Social skills and competences; and
- Thinking skills and competences.

The skills information from the ESCO framework is drawn from the European Dictionary of Skills and Competences (DISCO) which is an online thesaurus of over 100,000 terms, and approximately 36,000 phrases available in 11 European languages (www.disco-tools.eu). The purpose of DISCO is to support "the transparency and comparability of skills and competences by providing a well-structured and peer-reviewed multilingual terminology" (Müller-Riedlhuber & Ziegler, 2012). DISCO's categories of skills and competences include personal attitudes, values, and behavioral patterns and the thesaurus is based upon skill taxonomies from Austria, Germany, Sweden and the United States.

The skill taxonomy from the United States upon which DISCO and the ESCO are partially built comes from an initiative known as the Occupational Information Network (O*NET; www.onetcenter.org) sponsored from that country's Department of Labor. O*NET divides skills into a “basic” category that includes 10 skills which facilitate learning and a “cross-functional” category that includes 25 skills which facilitate performance of activities that occur across different groups of jobs. Within O*NET’s skills taxonomy, basic skills are either defined as content (e.g., mathematics) or process skills (e.g., critical thinking) while cross-functional skills are divided into social (e.g., negotiation), technical (e.g. programming), systems (e.g., systems analysis), and resource-management (e.g., time management) categories in addition to the non-categorized skill of complex problem solving.

Beyond more traditional research-based taxonomies, skill levels, there are a number of innovative methods being undertaken by the private sector, the public sector, and civil society. One method being utilized by stakeholders is to analyze publicly available and/or proprietary information on job postings, job descriptions, and self-reported information about skills from job incumbents and people seeking jobs. These efforts are being undertaken by a variety of stakeholders including private companies like Burning Glass, LinkedIn, and Monster. In addition, a tool supported by DISCO is the Europass website, specifically the Europass Curriculum Vitae and Skills Passport features (www.europass.cedefop.europa.eu). These features of the Europass site allow users to create standardized, geographically and linguistically portable, and automated profiles of their skills and qualifications via self-assessment and the recording of past experience. According to CEDFEOP (2014), in 2013 the Europass website registered 20.85 million visits and generated 10.16 million CVs and 1.3 million European skill passports.

Beyond taxonomies, international institutions, national governments, and individual stakeholders and coalitions of stakeholders from academia, the private sector, and civil-society have conducted reviews and/or studies to identify skills which are likely to be important for success in education, training, and the world of work in the 21st century. These efforts include an international survey and review by the OCED (2009) and an international effort led by the University of Melbourne in collaboration with the country governments of Australia, Costa Rica, Finland, the Netherlands, Singapore, and the United States and sponsored by Cisco, Intel, and Microsoft corporations (www.atc21s.org). In a report for the United States National Research Council, Pellegrino and Hilton (2012) reviewed and synthesized some of these and other efforts inclusive of academic research from cognitive, developmental, educational, organizational, and social psychology and economics. In general, the findings noted considerable cross-disciplinary confusion regarding: the conceptualization and characteristics of, appropriate forms of measurement for, effective methods of improving, and educational and workplace outcomes of obtaining 21st century skills. Despite this confusion, the report produced a taxonomy designed to align purported 21st century skills from several prominent reviews (Table 1). Pellegrino and Hilton (2012) used the term “competencies” to indicate the frequent combination of knowledge from educational and working contexts (e.g., from academic fields and industries) with skills that generalize across those contexts. This definition of competencies is in rough alignment with this working paper’s working definition of skills which includes consideration of such context; therefore, in Table 1, we use the word skills instead of competencies.

Table 1. Categories, clusters, and examples of 21st century skills from Pellegrino & Hilton (2012)

Category	Skill cluster	Example(s)
Cognitive	Cognitive processes and strategies	Critical thinking Argumentation
	Knowledge	Written communication Information and communication technology literacy
	Creativity	Idea generation
Intrapersonal	Intellectual openness	Cultural awareness Intellectual interest
	Conscientiousness	Metacognition Professionalism
	Positive core self-evaluation	Self-monitoring
Interpersonal	Teamwork and collaboration	Conflict resolution Negotiation
	Leadership	Assertive communication Self-presentation

2.4 Advanced approaches to measuring skills in international development

Historically and to a great extent today, proxies for skills (e.g., years of education, wages, and tools/technologies) have often been used to understand the role of skills within society. This approach to measurement differs from methods often utilized in education and psychology to estimate and measure skills as underlying "latent" psychological constructs that are reflected by certain observable behaviors or responses to surveys. In an effort to better understand and measure skills, major stakeholders in the international development community have adopted psychological approaches to skill measurement – for example, in studies of skills by the OECD (2013) and World Bank (2014b). The above efforts were multinational surveys in OECD and select lower-income countries respectively. The OECD's (2013) study known as the Survey of Adult Skills (PIAAC) evaluated proficiency in literacy, numeracy, problem-solving, and ICT skills. These evaluations were in part carried out by utilizing advanced psychometric approaches based on item response theory. The World Bank's (2014b) study, still underway, focuses on measuring generic skills of the working-age population in select lower-income countries. The effort, known as the Skills Towards Employment and Productivity (STEP) survey is comprised of both household-based and employer-based surveys. As a major innovation for major attempts at skills measurement in international development, the skill assessments employed in the STEP survey included both self-report measures of certain skills (e.g., written passage comprehension) and complementary testing of proficiency in those skills. Future efforts to evaluate skills will need to continue and build upon the methodological innovations.

3. Current uses of data-intensive science at work

While not often undertaken in an international development context, various social science fields prominently including education, I-O psychology, and vocational psychology have utilized information technologies and data-intensive scientific methods to both understand and develop people's skills in TVET and workplace contexts. We refer jointly to these efforts as data-intensive work-related psychology and education. More recently, initiatives emerging from a diversity of disciplines engaging with advanced information technologies (e.g., computer science and information-systems studies) have begun to engage with the world of work – particularly in the management and human-resources domains. Some of the work related to human resource issues has been labeled as "workforce science." Similarly, initiatives emerging alongside the widespread use of advanced communication technologies (e.g., mobile phones) have begun to engage with education – leading to interdisciplinary efforts with influences from such disciplines as education and communication studies known as "e-education." In this section, we review and highlight some major developments outside of the context of human development to understand their relevance and potential inside that context.

3.1 Data-intensive approaches from vocational and I-O psychology

Until recently, challenges to data acquisition, management, and analysis in relation to skills and broader human-resource issues has been presented less by large volumes of data, and more by issues of data velocity, variety, and veracity. To date, perhaps the primary areas of overlap between the issue of skills and data-intensive science come from: (i) the integration of computer-adaptive tools and high-fidelity simulations into education and human-resource management, (ii) the development of advanced analytical techniques to aggregate and connect multiple datasets that measure psychological concepts and vary in their veracity, and (iii) the use of advanced analytics to explore, model, and understand psychological constructs like skills. Often, these data-intensive approaches are tied to the implementation of information and communication technologies in the workplace. In their edited book, "The psychology of workplace technology," Coovert and Foster Thompson (2013) profiled much of the overlap between technology and the workplace and gathered expert contributors who discussed issues from technology-based hiring, performance appraisal, and teamwork to robots and social media.

3.1.1 Computer-adaptive solutions

In the fields of education and human resources, computer adaptive-approaches to psychological evaluation and ability assessment began in the 1980s; moreover, the advent of computer-adaptive training can be traced back to the 1970s with the development of precursors to intelligent tutoring systems (Ford & Meyer, 2013; Wainer, 2000). In tandem with advances in item response theory (IRT; mentioned in the previous section), greater computer processing capabilities have allowed for the widespread adoption of computer-adaptive testing and training to evaluate psychological constructs including knowledge, skills, abilities, and other personal characteristics like personality traits (known as KSAOs).

Computer-adaptive testing holds a number of features and advantages over classical testing approaches, including:

- Customization of test difficulty to users in a continuous and dynamic fashion;
- The more accurate estimation of an examinee's underlying ability with fewer test items;
- Full computer automation of tests and immediate scoring;
- Freedom for test takers to work at their own pace; and
- Greater test security.

Computer adaptive tests process the responses of test takers to each individual item and then use customized algorithms designed to predict the test-taker's ability level and customize the next item to more precisely measure that ability level. Within a human-resource context, computer adaptive tests can be used to evaluate a candidate's KSAOs for the purposes of vocational interest assessment, recruitment, hiring, training, and career development.

Computer-adaptive training also holds a number of advantages over non-adaptive training approaches, including (see Salas, Tannenbaum, Kraiger, & Smith-Jentsch, 2012):

- Increased training effectiveness and speed;
- The automated delivery of real-time feedback;
- Customization of training content and methods to trainee ability levels and training needs; and
- The fuller integration of virtual/simulated training environments.

One major development within the realm of computer-adaptive training has been the growing use of games complete with simulated environments. As with computer-adaptive tests, computer-adaptive training – especially in virtual and game environments – requires managing and analyzing real-time data generated by trainees and then using that information to inform the course of the training and/or provide interpretable real-time feedback. This process presents major challenges to the management and analysis of data.

The continual advance in ICT speed and prevalence has allowed computer-adaptive testing and training to grow dramatically in terms of its complexity and scope. For example, the benefits of computer-adaptive testing can simultaneously be made accessible to a seemingly unlimited number of people through un-proctored Internet-based testing. In addition, thanks to network technologies from mobile phones to web-based virtual environments, computer-adaptive training programs can facilitate the training of teams in geographically distributed locations.

3.1.2 Integrating varied datasets:

A second example of data-intensive science that has existed in the context of skills is the use of analytic methods within education, vocational psychology, and I-O psychology to analyze data from a variety of sources and of a variety of different types. While often not large enough to require non-traditional acquisition or management, advanced meta-analytic and multivariate methods are often necessary to compare and synthesize data from sometimes hundreds or thousands of studies that include different reliabilities, sample characteristics, models, and relationships. Advanced analytical strategies can include the generation of comparable metrics (e.g., effect sizes), the use of differential weighting, and the inclusion of both fixed and random effects on multiple levels of analysis to account for the nested and variable structure of data, and the accounting for differential reliabilities within data sub-sets.

The practical implementation of big data analytical methods in relation to skills and broader human-resource management often occurs in pursuit of the analysis and management of performance at work. For example, relatively popular forms of performance appraisal are 360° systems which integrate data from multiple sources – often from sources with divergent response patterns and data with different characteristics like reliability. Advanced data acquisition, management, and analysis methods are often necessary in the applied and real-time implementation of such systems – especially in the increasingly popular use of multisource electronic performance monitoring systems that might consider everything from keystrokes to the number of posts used in online communities.

3.2 Increased data volume in human resources – the rise of workforce science

The emergence of “workforce science” (sometimes also referred to as human-resource analytics) has corresponded with an increased consideration of data with volumes that present additional challenges to traditional big data acquisition, management, and analysis. A distinctive aspect of workforce science is its interdisciplinary nature – involving stakeholders, theories, and research methods from: computer science, economics, business management, and legal studies (see e.g., Searle Center, 2014). As in other applied fields, workforce science stakeholders in academia have sometimes partnered with for-profit efforts. As one example, the Workforce Science Project at Northwestern University has partnered with Cornerstone OnDemand, a “cloud-based talent management” firm. As another example, Mercer - a human resource consulting firm which has developed an institute devoted to workforce science – partnered with the World Economic Forum to publish a “Human Capital Report” which included global measures of the strength of over 120 countries’ workforces (World Economic Forum, 2013). As noted in that report, while considering worker skills was of central importance to this effort “there are no standard, internationally comparable datasets that directly measure skills, talent and experience despite agreement among governments, academia and business leaders that these should be measured. Therefore, the [the report] relies on a number of proxy variables…” (p. 5). Proxy variables considered in the report include: scientific and technical journal articles per capita, median age of the working population, and firm-level technology adoption. The measures considered in the World Economic Forum report are examples of the types of data that are often considered by workforce science.

While the field is highly disciplinary and still emerging making its exact characteristics unclear, it does seem that workforce science has concentrated on drawing insight from biographical (e.g., previous employment patterns), behavioral (e.g., number of computer keystrokes), and contextual data (e.g., temperature of office environments) that human-resource departments have chosen not to analyze, been unable to effectively mine for insight, or been unable to consider in combination. While various types of data being considered in the workforce science arena are relatively new to disciplines that have long engaged with human-resource issues, a large share of data currently focused on by workforce science practitioners, including biographical and behavioral data, have a long history within the field of I-O psychology (see Farr & Tippins, 2010).

3.3 Increased uptake of communication technologies – the rise of e-education

With the increasing advent of advanced communication technologies – especially Internet-based modes of education – the term "e-learning" has at times been adopted to refer to forms of skills development facilitated by ICTs (UNESCO, 2007). Massively open online courses (MOOCs) are a prominent example of e-learning. The operation of MOOCs often involve the use of big data management and analysis due to the enrollment of large number of course takers, the complex interactions of MOOC course takers in an online environment over a sustained amount of time, and the requirement for some MOOC in delivering real-time interaction and feedback (for example, grades to quizzes). Studying MOOCs is an important way to explore the potential of these e-learning platforms to both evaluate and improve skills as MOOCs have shown promise for educating large numbers of people at lower costs than traditional education (Griffiths, Chingos, Mulhern, & Spies, 2014). Despite this potential, MOOCs have sometimes failed to translate initial interest into course participation and completion with high rates of attrition at several stages along the pathway (e.g., Amnueypornsakul, Bhat, & Chinprutthiwong, 2014). Moreover, MOOCs have also at times failed to translate course success into important outcomes like employment (Kolowich, 2013). Regardless, major MOOC platforms, prominently including Udacity, seem to be transitioning away from broad-based education to more targeted vocational training (Adams, 2013).

Despite their potential to be relevant to human development, there is evidence that the skill-building potential of MOOCs is being utilized overwhelmingly by well-educated and wealthy people in both higher- and lower-income countries around the world (Emanuel, 2013). Evaluating the skill-building potential of MOOCs, especially for disadvantaged populations, will require a better understanding of a range of issues, including: (i) why people engage in MOOCs, (ii) what factors lead to course participation and completion, (iii) whether course participation and completion leads to skill building, and (iv) whether the skills that are built facilitate other important outcomes including employment and income growth. Research has begun to explore some of these issues and has often required data-intensive retrieval and data-intensive analysis (see for example, Rose et al., 2014). However, as argued by Wise (2014), considering such psychological concepts as learning activity requires the integration of theoretical models alongside the collection of large volumes of data regarding course taker behaviors.

4. Current uses of data-intensive science in international development

As highlighted by the concept note preceding this working paper (see Appendix A) and elsewhere (e.g., ITU, 2014; UN Global Pulse, 2013; World Bank, 2012c; World Bank, 2014a) there have been a number of initiatives that have sought to apply ICTs and data-intensive science to issues of global human development. These initiatives are in response to a rapid and global expansion of ICTs. By the end of 2014, there will be 3 billion Internet users – two-thirds of them coming from the developing world (ITU, 2014). However, the application of ICTs and big data to development is not merely a solution looking for a problem; human capabilities and development are inextricably linked with trends in information and communication created by an increasingly networked world of work and knowledge-based economy. In general, many of these initiatives are responding to the unprecedented spread of communication technologies in lower-income settings and the concomitant creation of a large volume of often real-time data through information technologies. As seen in Figure 8, the use of data-intensive science and ICTs for human development can be mapped by placing related but distinguishable fields and topics inside the domains of both human development and ICTs and arranging them in reference to their use of predominately information- or communication-based technologies.

In particular, Figure 10 includes the interdisciplinary field of information and communication technologies for development (ICTD) which considers ways in which ICTs can be used to assist human development both inside and outside of the world of work and TVET. ICTD practitioners also frequently engage with consideration of various psychological issues (Behrend, Gloss, Howardson, Thompson, & McCallum, 2013). While the field of ICTD involves a wide range of data-intensive approaches facilitated by advanced information technologies, it is often focused most explicitly on the potential of the "communicative" aspects of ICTs to assist human and economic development. In contrast, a diverse set of stakeholders has begun to focus more overtly on the potential for information technologies and data-intensive science to assist development in general and both economic development and workforce development in closer relation to the world of work and TVET. In this section, we review developments within the field of ICTD and then review emerging efforts for data-intensive development efforts in a human development context.

4.1 Information and communication technologies for development (ICTD)

The emergence of the ICTD field and much of the application of data-intensive science to development has been made possible by the rapid spread of ICTs across the world and into lower-income settings. Combining insights from the information and communication sciences, computer science, the organizational/management sciences, sociology, and development studies, ICTD has focused largely on promoting economic development and empowerment through research and applied interventions on both organizational and country levels of analysis (Gomez, Baron, & Fiore-Silfvast, 2012; Heeks, 2010). The ICTD field emerged in relation to the study of telecenters in lower-income settings and today is focused in large part on the potential for mobile phones and other communication technologies to enhance human development (Gomez et al. 2012). Common interventions in the field include the use of mobile phone text-message reminders to assist the performance of community health workers in

Tanzania (e.g. DeRenzi et al., 2012), the testing of a phone-based job application system in India (White et al., 2012), and the introduction of a mobile-phone based computerized decision support system to assist healthcare workers in Kenya (Anokwa et al., 2012).

As reflected in its frequent inclusion of an organization-level of analysis (many times including schools and work organizations), the ICTD field frequently considers issues that relate to the world of work and TVET. An example of an ICTD project that relates to the world of work and TVET is captured in a study by Cai, Abbott, and Bwambale (2013). That study explored the use of videos in the training of farmers in rural Uganda and found that using a combination of in-person and video-based lectures holds the potential to reduce pre-existing knowledge gaps between men and women. While the study by Cai et al. (2013) involved traditional research approaches, data-intensive approaches to studies and applied interventions are commonplace in the field of ICTD. As an example of a data-intensive ICTD study, Reda et al. (2012) utilized the scripting languages Pig and Hive on top of Hadoop (common tools for data-intensive science) to aggregate data from LinkedIn user profiles and user activity and then describe the growth, adoption and characteristics of social networking usage in lower-income countries. Insights from this study include the observations that the local geographic centrality /interconnectedness of LinkedIn users is lowest in Africa and the Middle East and that country membership in LinkedIn is not representative of the average educational attainment of those countries as LinkedIn users have nearly uniformly high levels of educational attainment.

As mentioned previously, the ICTD field also exhibits a focus on various psychological issues within the world of work and issues relating to TVET. Behrend et al. (2013a) conducted a cross-case analysis of proceedings from a major ICTD conference and identified that ICTD projects and studies frequently consider work-related psychological dynamics including psychological health and wellbeing, leadership, organizational culture, and work motivation; moreover, their findings indicate that ICTD interventions might frequently be disrupted by problems relating to these psychological dynamics (e.g., poor leadership). The importance of psychological dynamics to the success or failure of ICTD-related interventions is supported by another review, this time of articles published in the journal *Information Technologies & International Development* (Dodson, Sterling, & Bennett, 2012). In this review, it is noted that psychological dynamics like skills and thought patterns are important “deviation amplifying feedbacks” that are difficult to predict and that can disrupt the intended impact of ICT interventions (p. 57). Drawing lessons from the reviewed studies, Dodson et al. (2012) note that in order to support the success of ICTD interventions, greater attention needs to be placed on the “psychological, economic and socio-cultural factors that need to be in place at the start of a development initiative” (p. 61). More specifically to individual ICTD interventions, Behrend, Gloss, & Thompson (2013) reviewed various ICTD case-studies and noted strong interrelationships of the underlying goals, methods, and topical foci of I-O psychology and the ICTD field.

Beyond its consideration of psychological workplace dynamics, the field of ICTD has paid particular attention to the issue of skills. Indeed, the 2013 *International Conference on Information and Communications Technologies and Development* featured an entire session devoted to papers relating to skills (see, e.g., Molapo & Marsden, 2013). The focus on skills within the field of ICTD includes the skills of those attempting to undertake an intervention, the skills of a population as an influence on the success of an intervention, and skills as a desired end-state or outcome.

4.2 Data-intensive development efforts

As highlighted by the International Telecommunications Union (2014), the United Nations (2013), and the World Bank (2014a), data-intensive science holds great potential for promoting human development. As discussed in the previous section, a large share of this potential exists within the field of ICTD. However, a diverse range of related but differential efforts exist outside of ICTD that while still often involving communications technologies (e.g., mobile phones), prioritize the potential of information technologies. This point is reflected in a recent report from the Gates Foundation (2014) that instead of focusing on the role of mobile *phones* for development, focuses on using mobile *data* for development. The focus on utilizing mobile-phones and mobile-phone data (e.g., World Bank, 2012c) in the context of development is not accidental. By the end of 2014, mobile-cellular penetration in developing countries will reach 90% (ITU, 2014). This level of global penetration represents a tremendous expansion of available data that is in touch with lower-income populations. Indeed, as noted by the Gates Foundation (2014) report, mobile devices generate information about a user's "identity, location, social patterns, movement, finances and even ambient environmental conditions" from a diversity of mobile network sub-systems including mobile phone handsets, call detail records, prepaid billing services, customer relationship services, and value-added services which raises tremendous potential for this information to be put ot productive use for major issues in human development including economic development, health, and agriculture (p. 4). Despite this potential, important concerns exist regarding privacy, cybersecurity, and the protection of potentially secure information (e.g., financial or health records; ITU, 2013).

More generally than just data from mobile phones, and as mention by the World Bank (2014a), data-intensive science in an international development context often involves one or both of two approaches, namely: (i) utilizing data outside of an organizational context in order to heighten awareness or inform decision-makers, or (ii) utilizing data from within an organization to streamline and/or improve services. An example of a stakeholder who utilizes data-intensive science outside of an organizational context is the UN Global Pulse (2012) initiative with has labs devoted to the use of data-intensive science for development in Jakarta, Kampala and New York. Examples of stakeholders who utilize data-intensive approaches to streamline and/or improve services include a wide number of corporations operating the world over who are defined by ICTs (e.g., Amazon, Facebook, Google, IBM, Microsoft) and corporations who rely on ICTs to accomplish aspects of their missions (e.g., logistics or human resources management) such as Hyatt and Nissan. In addition, increasingly a diverse array of consulting firms devoted in large part to data-intensive business solutions including Visier, Cornerstone OnDemand, and Mattermark while prominent well-established firms like Mercer, McKinsey, SAS, SAP, SHL, and Oracle have begun to emphasize their own data-intensive business solutions. In connection to human development, many of these firms have partnered with not-for-profit organizations and multilaterals to engage in projects for a social or common economic good. For example, SAS has assisted the UN Global Pulse lab in a project to understand and predict economic trends in society through social media (SAS, 2011).

Prominent international multilaterals involved in the production, monitoring, and use of statistics in relation to data-intensive science include the International Telecommunications Union (ITU, 2014) and the United Nations Statistics Division. In addition, a wide number of multilaterals and actors in civil society have produced one or more reports devoted to issues relating to “big data,” including The Aspen Institute (2010), the OECD (2013), TechAmerica Foundation (2014), the World Bank (2014a), and the World Economic Forum (2014). A variety of national-level public sector actors have contributed to the conversation – for example the White House (2014) released a report on big data and representatives from countries as far afield as Jamaica, South Africa, the Netherlands, and South Korea made presentations at a 2013 United Nations sessions on “Big Data for Policy, Development and Official Statistics” (UN Stats, 2013).

5. Data-intensive science to build human capabilities at work?

The previous four sections of this working paper have separately considered the potential for and conceptual overlap involved in the synthesis of data-intensive science, skills, the world of work and TVET, human development, psychology, and information and communication technologies. In this section, we bring all of these issues together and demonstrate the unique potential of what are naming information and communication technologies for human capabilities (ICT4H).

As described in Section 4, actors in international development and the field of ICTD have demonstrated the potential to enhance human development in lower-income settings, with marginalized populations, and in the informal economy. In Section 3, both relatively new and longstanding data-intensive scientific approaches to enhancing skills-development in predominately higher-income settings and within the realms of human-resource development and TVET were presented. In Section 2, the psychological influences and underpinnings of skills development were illuminated along with the connections between skills-development and the broader development of human capabilities. In Section 1, several ways in which information technologies have made possible and continue to facilitate big data acquisition, management, and analysis were detailed. In summary of both the above points and broad themes in the previous sections, there are strong interconnections among data-intensive science, information technology, communication technology, psychology, 21st century skills, and human development as conceptualized through the human capabilities approach. These interconnections help to create the unique potential for combinations of data-intensive approaches with ICTs to effectively measure and promote skills- and human capabilities for marginalized populations and people living and/or working in lower-income settings and the informal economy.

5.1 ICT4H activities

In Section 4, it was mentioned that there are two predominate purposes for data-intensive science in a human-development context, namely: utilizing data outside of an organizational context for policymaking/awareness and utilizing data from inside an organizational context to streamline or improve services. In ICT4H, work and educational organizations, institutions, and electronic platforms are relevant not just because they might run more efficiently based off of a successful project or better information, but also because they are major contexts in which skills development takes place. In addition, in contrast from other data-intensive development efforts, ICT4H includes salient goals specific to its unique consideration of psychology and the skill-development challenges in lower-income settings, the informal economy, and with marginalized populations, including:

(i) Promoting a better understanding of skills' role in society – including various facets of an overall skills conceptual framework, namely: skill acquisition, skill requirements, skill development processes in the workplace and/or TVET, the outcomes of skill development/utilization, and the contextual workplace and societal influences on these aspects of this framework;

(ii) Using data to enhance skill development efforts – either in a real-time or day-to-day system (e.g., an online training platform) or in providing feedback to human trainers for summative/formative feedback or to policymakers for policy guidance and program evaluation criteria;

(iii) Overcoming skill-coordination failures (e.g., the inefficient matching of job applicants with skill needs in a labor market) by providing either immediate feedback into information systems that help to evaluate the available skills of individuals and the skill needs of employers, or by providing managers or economic/workforce development practitioners with detailed information; and

(iv) Directly measuring/enhancing human capabilities by helping individuals to learn about themselves and their personal goals in relation to work (e.g., through self-administered skill assessments or personality/vocational-interest surveys), by providing decision support in relation to their goals (e.g., by providing relevant labor-market information and information about required skills within certain occupations), by enhancing individuals' skills and commitment to self and environmental exploration, and assisting in the removal of psychological barriers to skills development (e.g., through self-efficacy training).

(v) Providing aggregate data on the skills, interests, and goals of individuals and communities, which can be used by governments, non-profits, and the private sector in deciding how and where to invest to create new enterprises and new jobs.

5.2 Sources and characteristics of ICT4H data

Implicit in discussions from Sections 1 and 2 is the reality that communication technologies provide unique opportunities for the generation of data relevant to both important 21st century skills and human development as conceptualized through the human capabilities approach. Various cognitive, intrapersonal, and interpersonal 21st century skills are in large part both subjective and/or psychological in nature (e.g., the skill of self-monitoring). In addition, because human capabilities are ultimately expressions of both one's *perceived* and actual ability to pursue self-set goals, the measurement of human capabilities is at least partially and necessarily psychological. As stated in Section 1, sources of big data often emerge from one or more of four sources:

(i) The interface between a physical environment and technological devices (e.g., temperature sensors);

(ii) Digital connections between technological devices (e.g., between motor vehicles and mobile phones);

(iii) The interface between people and technological devices (e.g., mobile phones); and

(iv) Interactions among individuals in a digital environment (e.g., between people in a social network).

While biographical, behavioral, and/or environmental indices of education, health, and income can help provide evidence for the likelihood for individuals having both the perceived and actual ability to accomplish their goals, as discussed in Section 2, a broad range of psychological and non-psychological factors in the workplace, community, and broader society can intervene to create both psychological and non-psychological barriers to skills and capabilities. Therefore, even though useful data in relation to skills and human development arises from all four of the sources above, it seems likely that the most valuable, direct, and reliable data for the purposes of human- and skills- development will emerge from people's interactions with devices (e.g., mobile phones) and in digital environments with one another where cognitive processes, intrapersonal skills, interpersonal skills can be transformed into data, acquired, managed, and analyzed. However, even in digital environments and in person-to-device interaction, reliable data on certain psychological phenomena are difficult to obtain – for example, the degree to which a person feel able to learn a skill through education/training opportunities (known as training self-efficacy which is a critically important determinant of various aspects of training success). Establishing a person's training self-efficacy, or training self-efficacy trends for millions of people, could be accomplished through:

(i) Existing data that is passively generated (e.g., through monitoring of behavioral activity leading up to a training session);

(ii) Data that is generated through an active intervention on behalf of the investigator/analyst (e.g., the posting of a survey link on a company intranet); or through

(iii) Data "mashups" - combination of multiple forms of data from each one of these categories or a combination of different forms.

As mentioned in Section 1 during the discussion of challenges to traditional data acquisition, management, and analysis methods, one challenge emerges from low, variable, or unknown levels of data veracity – or reliability. As later discussed in Section 3, psychological latent constructs like skills can present challenges to quickly and easily estimating and accounting for variable levels of reliability in measurement. In light of this difficulty, an important distinction for ICT4H activities is whether they aim and/or are able to "directly" measure a construct of interest – that is, being able to utilize a highly reliable method of measurement (e.g., asking someone directly about whether they are satisfied in their job) or whether an "indirect" measurement approach that relies upon a meaningful degree of relatively unreliable inference is required (e.g., inferring satisfaction based upon physical activity).

5.3 Brief examples of potential ICT4H projects

To illustrate the unique potential and characteristics of ICT4H interventions and approaches, 10 different examples of ICT4H projects are provided in Table 2. Each example project is categorized according to the data types, purposes, link to skills (e.g., direct/indirect), and activities as listed above. These examples are merely meant to provide a sample of the breadth of potential ICT4H activities.

***Table 2**: Examples of potential information and communication technologies for human-capabilities projects (ICT4H)*		***Data type***			***Person to device data-source***				***Person to person data- source***				***Link***		***Activity***			
	Example	**Existing**	**To be created**	**Mashup**	**Search engine**	**Online survey/games**	**Job/employment website**	**Business/industry websites**	**Mobile phones**	**Social network**	**Professional network**	**E-mail**	**Direct/reliable**	**Indirect/inferred/unreliable**	**Understanding skills**	**Enhancing skill development**	**Overcome coordin. failures**	**Enhance/measure capability**
1	Use registry of phone number types (e.g. businesses vs. residential) and patterns of mobile calls to identify potential and existing entrepreneurs for interventions, e.g., entrepreneurial skills training	X							X					X	X	X	X	X
2	Use data on word and vocabulary complexity in mobile phone text messages to measure writing skills of disadvantaged populations	X							X				X		X			
3	Use descriptions of activities from company websites (e.g., services provided) in a given area to estimate regional skill requirements	X						X						X	X		X	
4	Content analyze Tweets for insight into discussions about skills prior to and following discussions regarding loss of job for insight into the demand for, role of, and/or lack of key skills	X								X				X	X		X	
5	Use computer-adaptive testing to create 1-minute mobile-phone math games for disadvantaged youth customized to respondents' ability levels, use info to build skills and inform curricula weaknesses		X			X							X		X	X		
6	Push a survey out over a mobile network to ask about individuals' 1-year skill-related goals: experiment with the effectiveness of providing reminders of the goal at different intervals by testing with a control group		X						X				X			X		X
7	Publish a fun "personal interest" survey via mobile phone (e.g., "What is your dream career?") that evaluates key transferrable skills and makes vocational recommendations complete with directions to training resources		X						X				X		X		X	X
8	Build a mobile-phone based employer toolkit for small businesses which allows employers to assess skill acquisition before and after training		X						X				X		X	X		
9	Create a jobs wiki which allows users from around the world to create and craft localized descriptions of jobs and the skills those jobs require		X				X							X	X		X	
10	Detect spelling mistakes or violations of social norms on LinkedIn by disadvantaged populations, provide access to a professional-skills training course, track job status changes	X	X	X							X		X			X		

5.4 In-depth example of ICT4H: Entrepreneurship

To more fully explore the unique potential and characteristics of ICT4H approaches, we review two key topics in the ICT4H domain, review existing ICT4H-related approaches, and then discuss ways to enhance those approaches to maximize their effectiveness in promoting human capabilities.

In higher-income settings, entrepreneurship is critical for job creation and economic growth; entrepreneurship creates innovation in any number of realms within society including art, commerce, and technology, and when otherwise unemployed – becoming self-employed as an entrepreneur can be a powerful way to enter the positive work cycle of setting goals, developing skills, and developing a sense of psychological empowerment. Due to higher levels of poverty, physical insecurity, and low-skill traps that tend to inhibit innovation, the need for jobs, economic growth, innovation, and psychological empowerment can be even greater in lower-income settings and the informal economy than in higher-income ones.

5.4.1 A data-intensive approach to understanding entrepreneurship

Outside of developing contexts, significant media attention in relation to entrepreneurship has recently been focused on a joint effort undertaken by the data-intensive start-up Mattermark and its client BloombergBeta. This effort involved the monitoring of publicly available information (e.g., job history and age) for people near San Francisco in the United States for a period of 9 months. Mattermark used this information to map out the most promising factors that predicted people's decisions to start a business. Using this mapping, the company then identified 350 likely future entrepreneurs based upon publicly available information about them on social networks (Griffith, 2014; Rose, 2014). According to its website, Mattermark claims that by accounting for age, education, profession, past job history, and various unspecific factors, its algorithm is 25 times better than chance in identifying future business founders (www.mattermark.com). The example of BloombergBeta and Mattermark's efforts to predict the emergence of entrepreneurs is a good demonstration of the characteristics of data-intensive approaches to developing profiles and making predictions in the emerging workforce science domain. Yet, the population, available data, and ICT-adoption of individuals near San Francisco are quite different from conditions in many lower-income settings and for disadvantaged populations even in the United States.

5.4.2 A psychological and data-intensive approach in lower-income settings

An effort that considers entrepreneurs in lower-income contexts, utilizes data-intensive methods, and also integrates theory and measurement approaches from psychology comes from the Entrepreneurial Finance Lab (EFL; www.efl.org). As stated on its webpage, EFL claims to have pioneered "psychometric credit scoring" which has been utilized to help identify high-potential entrepreneurs, improve credit underwriting for small businesses, and credit scoring for consumers in countries including Ecuador and Zimbabwe through the measurement of psychological traits (relatively stable characteristics). As noted in Klinger, Khwaja, and Carpio (2013), the inspiration for EFL emerged from the problem that many small business entrepreneurs (over 300 million) in emerging markets want loans but cannot obtain them; many of these entrepreneurs might otherwise qualify for those loans if they had a formally established credit history and collateral. After reviewing research on the psychological predictors of successful entrepreneurship, instruments measuring key psychological traits were administered to 1,580 small business entrepreneurs applying for loans at banks in Peru, Kenya, Columbia, and South Africa (Klinger et al., 2013).

The results of Klinger and colleague's (2013) pilot study established multiple significant and meaningful relationships between the psychological traits measured and multiple outcome of interest (including default risk on the bank loan and profitability of the entrepreneur's firm). However, these relationships differed substantially between banks and country contexts. Because of the differential relationships observed, the nested nature of the data, the small sample size from some banks, and a large number of covariates, Klinger et al. (2013) adopted Bayesian methods to identify entrepreneurs who would be most likely to make full and on-time repayment of loans while operating profitable firms. In the terminology of this working paper, the authors adopted a non-traditional analytical strategy to account for the variable veracity of subsets of their data. However, the analytical challenges for the EFL did not stop with their pilot of a relatively small number of entrepreneurs. Instead, with the implementation of the EFL's psychometric credit scoring system on a Bayesian analytical foundation and in a computer-based administration format comes the ability (indeed the financial imperative) to create a learning system that automatically integrates a continuous flow of new psychometric and performance information back into the banks' systems to help calibrate the predictive model and to hedge limitations in instances in which new populations, contexts, or changing situations might affect predictions but are not accompanied by sufficient information necessary to understand those new dynamics.

5.4.3 A psychological, mobile-based, and data-intensive approach to skills development in lower-income settings

Much of the psychological research reviewed by Klinger and colleagues (2013) included findings that have been recently synthesized by Gielnik and Frese (2014). These later authors collected meta-analytic findings regarding individual-level influences on entrepreneurial success including the relatively strong and positive influences of various stable personality traits; however, this review also pinpointed a positive role of human capital (e.g., knowledge and skills), and positive contributions for certain strategy types such as those enacted by people with so called entrepreneurial orientations. In light of the relevance of potentially trainable skills to entrepreneurial success, we consider an additional way in which ICTs can be combined with data-intensive methods for entrepreneurial skills-development in lower-income settings.

Despite the importance of individual-level traits to entrepreneurial success, traits – like personality characteristics and underlying cognitive abilities – are largely resistant to outside manipulation. Luckily, as Gielnik and Frese (2013) observe in their consideration of entrepreneurship in lower-income settings, there are two predominate methods to assist entrepreneurial success. The first method is "top-down" and involves the alteration of systems and norms that might be impeding successful entrepreneurial behavior (e.g., onerous paperwork to start a new business). The second method is "bottom-up" and involves assisting entrepreneurs to make more adaptive decisions and strategies (e.g., pattern-based recognition of business opportunities - e.g., Baron, 2006), learn new facts (e.g., technical knowledge germane to an entrepreneurs' business), and/or develop successful entrepreneurial skills. Relevant entrepreneurial skills include many 21st century skills identified in Section 2 like negotiation and more entrepreneurially specific skills like business planning (Chell, 2013). In terms of approaches to assisting entrepreneurial skills, it has been observed that educational approaches to the building of these skills – for example through TVET – are even more effective in lower-income settings than they are in higher-income settings (Unger, Rauch, Frese, & Rosenbusch, 2011). However, some entrepreneurial skills are difficult to train outside of realistic work contexts and in lower-income settings and in the informal economy, the ability to provide structured education and training is often quite limited.

As highlighted prominently by a report from the Brookings Institution (West, 2012) entitled "How mobile technology is driving global entrepreneurship," mobile phones pose a particularly important opportunity to effectively deploy meaningful entrepreneurship and to harness such skill-development alongside data-intensive approaches. Frequently, entrepreneurs utilize mobile phones in relation to informal forms of skills development by connecting with support networks and exchanging insight and best-practices from other entrepreneurs. Such a support network highlights that factors affecting entrepreneurial success exist not only on societal and individual levels, but also on the level of the group and/or community (Baron & Henry, 2011). Indeed, social capital is related to the diversity, strength, cohesiveness of a network, an individual's embeddedness within that network, and the makeup of that network (from family members to suppliers; Cheraghi & Schott, 2014). Issues of social capital are particularly important in the process of beginning a business and highlight how the very idea of entrepreneurship often requires social and community contextualization (Welter, 2010). Yet as further expounded by Baron and Henry (2011), the benefit of utilizing social capital and supportive networks for positive entrepreneurial outcomes like sales growth relies in large part upon individuals' strong social skills. Thus, social skills seem like a particularly important skill set to target through data-intensive methods facilitated by mobile technologies.

In order to engage with entrepreneurial social skills on a socially impactful scale, it can be usefully observed that many social behaviors relating to the maintenance and building of social networks take place within the digital environment of mobile phone calls. As highlighted earlier by the Bill & Melinda Gates Foundation (2014) report on using mobile data for development, social network analyses of call records, call metadata and in the case of smart phones browser and app usage can help to determine social patterns – in particular subscribers with high relative social influence. At the very least, calling patterns can be observed across time in order to understand the cellular communication patterns of entrepreneurs with varying degrees of success, from different sectors, social backgrounds, and regions. The study of such cellular network traces have been analyzed by a number of stakeholders involved in such fields as ICTD. For example, Zheleva, Scmitt, Vigil, and Belding (2013), analyzed the community persistence in egocentric social graphs from the Orange telecommunications corporation's network in Cote d'Ivoire. A promising step to transforming insight from network analyses of entrepreneurs' mobile calls would be to form taxonomies of calling patterns that can be linked with certain stages of successful business development (if indeed any exist). The connections of these patterns to the behaviors and/or situations they are intended to represent should be tested through attempting to identify entrepreneurs in a cellular network based purely on their calling patterns. Despite the usefulness of insights into entrepreneurs' social networks for academic and policy-based purposes, the practical implementation of these insights to assist the skills-development of entrepreneurs might have important benefits for lower-income settings.

In order to harness academic or policy-related insight into entrepreneurial social networks for the purposes of skills development, the psychological measurement technique of mental models can be employed to convert an exemplar entrepreneur's social skills into a spatial and digital form. This form can then be compared to and integrated with a visualization of an exemplar entrepreneur's social network as represented by calling patterns and/or metadata. The use of mental models in a work context have been used to understand entrepreneurial decision-making (e.g., Zahra, Korri, & Yu, 2005) and forms of interpersonal interaction and collaboration like teamwork (see Lim & Klein, 2006). The visual representations of entrepreneurs' social networks and social skills can be used in skills development by adapting them for use in training scenarios – often involving the comparison of the trainee's mental models to those of an expert.

5.4.1 Summary of approaches

As highlighted above, the aim of promoting entrepreneurship has been approached in higher-income contexts from the emerging field of workforce science. Moreover, in order to help establish the creditworthiness of entrepreneurs in lower-income settings, the Entrepreneurial Finance Lab has employed combinations of psychological measurement and data-intensive science to identify high-potential entrepreneurs who might otherwise not receive financial support. Following on from these innovative efforts, we suggest additional steps that jointly utilize ICTs, data-intensive science, and psychology for the development of entrepreneurship skills. In particular, we advocate for the potential analysis of data from mobile phone networks to develop a detailed understanding of entrepreneurial social networks in lower-income settings. Moreover, in order to transform this understanding into practical forms of skills development, we propose that expert entrepreneurs' mental models for social interaction can be visualized, integrated along with visual representations of their social networks, and used in training scenarios.

Generalizing from the specific example of the issue of entrepreneurship, data-intensive science, information and communication technologies, and psychology hold an especially great potential to promote skills- and human-development in light of:

(i) the central role of work-related skills in broader human capabilities;
(ii) the unique ability of communication technologies to illuminate and digitize important cognitive, intrapersonal, and interpersonal skills; and
(iii) the unique ability of data-intensive approaches facilitated by information technologies to draw insight from forms of data that prevent traditional forms of acquisition, management, and analysis.

6. Challenges and considerations

The rise of data-intensive science has brought solutions from new perspectives to old problems. Especially with marginalized populations and within lower-income regions and the informal economy, these new approaches and perspectives have the potential to enhance human capabilities. At the same time, there is a risk that these new approaches will result in mistakes that have been made previously by stakeholders and disciplines already working to solve these problems. In admission of the importance of drawing lessons from fields related to the use of data-intensive science, ICTs, and psychology for skills- and human-development, we briefly review key considerations and challenges from TVET, the organizational sciences, international development, and various disciplines that have considered aspects of the "digital divide."

6.1 Best-practices and theory in international development

Stakeholders involved in promoting international and human development have established best-practices and theories in light of decades of experience. For example, within the field of ICTD it has been noted that "top-down, technology-centric, goal-diffuse approaches to ICTD contribute to unsatisfactory development results" (Dodson et al., 2012, p. 56). These and other important considerations have been usefully summarized by the OECD in its Paris Declaration for Aid Effectiveness and later conventions by focusing on four principles (OECD, 2014):

- Ownership: it is critical to ensure that the countries and people affected by development initiatives exhibit ownership over those initiatives;
- A focus on results: The monitoring of progress toward development goals needs to be both prioritized and it needs to involve embedding of data collection necessary to support that monitoring within existing stakeholder frameworks (e.g., those of countries);
- Partnerships: Inclusiveness in developmental efforts is critical to their long-term success and this prominently includes involving non-state actors including civil-society and private-sector actors;
- Transparency: Transparency regarding the effectiveness of efforts is important to avoid an abuse of power and waste of resources.

The incorporation of data-intensive science into development activities presents the potential to measure the results of developmental activities far more effectively than previously possible. At the same time, because the digital divide continues to exist through disparities in effective use of ICTs and the availability of ICT-related data, the possibility for developmental activities to sidestep the active ownership of local stakeholders and to lose inclusiveness and transparency is especially great. In its report, "New Data for Understanding the Human Condition" the OECD (2013) admits to these dangers and outlines a set of recommendations to help local stakeholders effectively utilize big data. These recommendations include (but are not limited to):

- Developing a code of conduct regarding data privacy and ethical research norms for data-intensive science;

- Establishing mechanisms for producers and users of data to facilitate their shared use and analysis;
- Prioritizing the harmonization of data collection, metadata, and data formats.

Because of the importance of semantic web technologies to the analysis and sharing of big data relating to skills (see Tippins & Hilton, 2010), the importance of developing a definitive international ontology for skills is of paramount importance. Of note, the HR Open Standards Consortium has developed standards for data on such subjects as performance management, recruiting, and screening of employees in a human-resource managed environment (www.hropenstandards.org). Moreover, stakeholders in the European Union continue to innovate in regard to the generation of skill ontologies through the DISCO project (www.disco-tools.eu).

As previously mentioned, some approaches to data-intensive science tend to concentrate on inductive methods of drawing insight from predominately behavioral, biographical, and environmental data. These approaches run the risk of deemphasizing or obscuring subjective and/or psychological aspects of skills- and human-development which might confuse or prevent an understanding of important 21st century skills. In addition, these approaches run the risk of obscuring subjective and psychological aspects of what we argue to be the ultimate human development criterion – namely human capabilities. While a number of behavioral and physical indicators can be innovatively applied to better understand human capabilities, ultimately the full measurement of capabilities emerges from a person's own expression of the degree to which they consider themselves capable of accomplishing the things they have reason to value in life.

One of the meta-themes from best-practices in international development is to establish highly cooperative, interdisciplinary, and cross-sector groups of stakeholders to undertake any given development initiative. While challenging, these groups are one of the best hopes for long-term sustainability and appropriateness of efforts. Such partnerships will be both especially critical and difficult in relation to ICT4H due to a myriad of major concerns and considerations relating to, among other things, the data-analytic skills of all stakeholders, privacy concerns, and data-security issues. Indeed, there are many other often unclear concerns and surprises in the rapidly developing world of data-intensive science.

6.2 Insight from TVET – challenges and priorities for integrating ICTs into TVET

As indicated by a report from the UNESCO International Centre for Technical and Vocational Education and Training (UNEVOC; 2013a), an important challenge relating to the growing use of data-intensive science and ICTs for the purposes of skills development is to ensure that those efforts are aligned to existing and emerging international TVET frameworks.

In addition to the above, UNEVOC (2013a) noted the importance for innovative ICT solutions to help improve both the quality and attractiveness of TVET as a skill-development pathway. In a related report entitled "Revisiting Global Trends in TVET: Reflections on Theory and Practice," (UNEVOC, 2013b) the case is made that improving the attractiveness of TVET is likely dependent on, among other things, greater demand from top employers and from youth and their families. One of the unique opportunities inherent in combining ICTs and data-intensive science with TVET is the potential to integrate TVET outcomes and processes with exciting and innovative developments in online skills recognition (including the recognition of both ICT-related skills like programming and skills more generally). For example, sources of interest and attention to this area include the prestige of new forms of recognition (e.g.,www.openbadges.org), winning online skills competitions (e.g.,www.topcoder.com), and the greater integration of professional and social networks into the world of work (e.g.,www.linkedin.com).

6.3 Lessons from organizational sciences – the dangers of ignoring theory

A potential tendency within data-intensive science is to rely upon inductive observation to the exclusion of deductive forms of theoretical reasoning. Empirical observation supported by an abundance of data on a large number of factors holds tremendous promise for both skills- and human-development. However, the avoidance of incorrect, counterproductive, and/or harmful conclusions and interventions regarding human behavior and cognition is reliant upon allowing those inductive insights to be informed by broader theoretical insights into education, skills development, and vocational/I-O psychology. In particular, theory is important because it helps researcher and practitioners to:

- Make valid, and helpful, causal inferences from empirical observation;
- Move beyond the data at hand to generate broader inferences; and
- Avoid the unreflective importation of potentially harmful biases.

We discuss each consideration in more detail below.

6.3.1 Causal inferences

As illustrated by the maxim, "correlation does not necessary imply causation," it is critical that data-intensive science utilizes methods to rule out alternative explanations to apparent signals that might imply misleading causal relationships. Data-intensive science has at its disposal a number of important approaches to ensuring its explanations and predictions are valid. These approaches including the ability to statistically control for a large number of potentially confounding variables and the ability to simulate the effects of potential interventions on a system before making recommendations. However, there is sometimes no replacement for the benefits of randomized experimentation (also known as A/B testing) in ruling out alternative explanations and ensuring the validity of causal inference (Shadish, Cook, & Campbell, 2002). Theory is critically important to help guide randomized experimentation because not all manipulations are likely to be appropriate and/or helpful and theory derived from past induction can help to focus researchers on accounting for dynamics that might otherwise not be included in an experiment – even in the world of large volumes of data.

6.3.2 Moving beyond the data at hand

Induction is by definition based exclusively on the data at hand and makes predictions based only on what has come before. However, it is not the case in the rapidly changing world of work that what has happened before will continue to happen in the future and that what works in one cultural and socioeconomic context will work in other contexts. Perhaps one of the greatest dangers to inductive reasoning is the inability to account for considerations and dynamics that exist outside of the contexts and populations in question (for example, the cultural values of people not included in a given organization or nation). The dynamic interplay between inductive observation and deductive theoretical sense-making is the best way to both ensure that inferences are grounded in empirical observation and consider or at least be aware of dynamics that exist outside of the frame of reference.

6.3.3 Avoiding biases

As has been argued in a recent editorial from a top academic business and human-resource management journals, the social and behavioral sciences have been littered with misguided and failed attempts to focus exclusively on data to the exclusion of theory (Suddaby, 2014). This phenomenon is often known as "dustbowl empiricism." Dustbowl empiricism and/or a broader adherence to a "value-free" or "theory-free" form of science or research is quite dangerous in the social sciences – especially when research is geared to help bright about socioeconomic change. As argued by some, including the ethicist and I-O psychologist Joel Lefkowitz (2008), an adherence to supposedly value or theory neutral science often can hide the researcher's implicit values and theories and makes them harder to reflect upon or question. Moreover, when such perspectives are claimed, it is the values, priorities, and perspectives of the status-quo in society that often remain intact.

6.4 The remaining digital divide

Despite the rapid increase in digital technologies in lower-income settings, the proverbial "digital divide" remains (ITU, 2014; van Deursen, Courtois, & van Dijk, 2014). The seemingly paradoxical growth of digital technologies in lower-income setting alongside a continuing digital divide can be explained in two ways. First, effective utilization of ICTs is dependent on more than simple access to those technologies. From a technological standpoint, not all ICTs are created equal and an individual's capability to communicate, collaborate, find information, and efficiently accomplish a range of tasks is heavily dependent on both the qualities of the technological device and both the reliability and data capacity of the digital connection. Moreover, acceptance and use of technology is dependent on people's expectations regarding (Venkatesh, Morries, Davis, & Davis, 2003):

- (i) the degree to which the technology will assist them;
- (ii) people's perceived ease of adopting and using technology; and
- (iii) people's perceptions regarding the social acceptability of technology use.

From a complementary perspective, effective utilization of digital technologies can be said to be dependent on people's knowledge of the benefit of digital opportunities, their skills in utilizing those opportunities, and their motivation to interface with those opportunities (Van Dijk, 2005). Thus, while ICTs and the use of data-intensive science present tremendous opportunities for promoting development goals, innovative approaches will need to be utilized to overcome frequently, if not predominately, psychologically-related barriers to the adoption and effective utilization of ICTs.

In addition to individual-level differences in the adoption and effective utilization of ICTs, the digital divide is reflected on a macro level of analysis in the dearth of data available to monitor and guide development policy and interventions in lower-income countries. As the United Nations (2014) illustrates in the 2014 Millennium Development Goals Report, a lack of basic data including the number and quality of jobs is missing in many lower-income countries. The OECD et al., (2013) reinforce this finding in relation to data on skills as they are often missing in low- and lower-middle income countries.

References

Adams, S. (2013, 12 November). Are MOOCs really a failure? *Forbes*. Retrieved from www.forbes.com

Amnueypornsakul, B., Bhat, S., & Chinprutthiwong, P. (2014). Predicting attrition along the aay: The UIUC model. *EMNLP,* 55-67.

Anokwa, Y., Ribeka, N., Parikh, T., Borriello, G., & Were, M. (2012). *Design of a clinical-decision support system for resource-limited settings*. Stories from the field: Paper presented at ICTD 2012: The International Conference on Information and Communication Technologies and Development, Atlanta, United States.

Aspen Institute. (2010). *The promise and peril of big data*. Retrieved from www.aspeninstitute.org

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1-26.

Baron, R. A. (2006, February). Opportunity recognition as pattern recognition: How entrepreneurs connect the dots to identify new business opportunities. *Academy of Management Perspectives*.

Baron, R. A., & Henry, R. A. (2011). Entrepreneurship: The genesis of organizations. .In S. Zedeck, (Ed.), *APA handbook of industrial and organizational psychology: Vol. 1 Building and developing the organization* (pp. 241-273). Washington, DC, United States: American Psychological Association.

Behrend, T. S., Gloss, A. E., Howardson, G., Thompson, L. F., & McCallum, S. (2013, December). *Psychological dynamics in ICTD projects*. Paper presented at ICTD 2013: The International Conference on Information and Communication Technologies and Development, Cape Town, South Africa.

Behrend, T. S., Gloss, A. E., & Thompson, L. F. (2013). Global development through the psychology of workplace technology. In M. D., Coovert, & L. F., Thompson (Eds.), *The psychology of workplace technology* (Society for Industrial and Organizational Psychology Organizational Frontiers Series; pp. 261-283). New York, NY: Routledge Academic.

Bischoff, K. M., Gielnik, M. M., & Frese, M. (2013). Entrepreneurship training in developing countries. In W. Reichman (Ed*.), Industrial and Organizational Psychology Serving the Underderserved*. New York, NY, USA: Palgrave Macmillan.

Blustein, D. L. (2006). *The psychology of working: A new perspective for career development, counseling, and public policy*. Mahway, NJ : Erlbaum.

Blustein, D. L., McWhirter, E. H., Perry, J. C. (2005). An emancipatory communitarian approach to vocational development theory, research, and practice. *The Counseling Psychologist, 33*(2), 141-179.

Cai, T., Abbott, E., & Bwambale, N. (2013). *The ability of video training to reduce agricultural knowledge gaps between men and women in rural Uganda*. Paper presented at ICTD 2013:

The International Conference on Information and Communication Technologies and Development, Cape Town, South Africa.

Chakroun, B., & Daniel, R. (2014). *Big Data for world of work exploring the possibilities* (Concept note). Paris, France: United Nations Educational, Scientific, and Cultural Organization.

Chell, E. (2013). Review of skill and the entrepreneurial process. *International Journal of Entrepreneurial Behavior & Research*, *19*(1), 6-31.

Chen, M., Mao, S., Zhang, Y., & Leung, V. C. M. (2014). *Big data: Related technologies, challenges and future prospects*. doi:0.1007/978-3-319-06245-7

Cheraghi, M., & Schott, T. (2014). *Size, diversity and components in the network around an entrepreneur*. In F. Can (Ed.), *State of the art applications of social network analysis*. doi: 10.1007/978-3-319-05912-9_16

Chernyshenko, O. S., Stark, S., & Drasgow, F. (2011). Individual differences: Their measurement and validity. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology: Vol. 2 Selecting and developing members for the organization* (pp. 117-151). Washington, DC, United States: American Psychological Association.

Coovert, M., & Foster Thompson, L. (Eds.). (2013). *The psychology of workplace technology* (Society for Industrial and Organizational Psychology Organizational Frontiers Series; pp. 261-283). New York, NY: Routledge Academic.

DeRenzi, B., Findlater, L., Payne, B., Mangilima, J., Parikh, T., Borrielo, G., & Lesh, N. (2012). *Improving community health worker performance through automated SMS*. Paper presented at ICTD 2012: The International Conference on Information and Communication Technologies and Development, Atlanta, United States.

Dodson, L. L., Sterling, S. R., and Bennett, J. K. 2012. *Considering failure: Eight years of ITID research*. Paper presented at ICTD 2012: The International Conference on Information and Communication Technologies and Development, Atlanta, United States.

Farr, J. L., & Tippins, N. T. (2010).*Handbook of employee selection*. New York, NY, USA: Routledge.

Ford, J. K., & Meyer, T. (2013). Advances in training technology: Meeting the workplace challenges of development, deep specialization, and collaborative learning. *The psychology of workplace technology* (Society for Industrial and Organizational Psychology Organizational Frontiers Series; pp. 43-76). New York, NY: Routledge Academic.

Frese, M. & Gielnik, M. M. (2014). The psychology of entrepreneurship. *Annual Review of Organizational Psychology and Organizational Behaviour*, *1*, 413-438.

Gagne, M. & Deci, P. (2014). The Oxford handbook of work engagement, motivation, and self-determination theory. New York, NY, United States: Oxford University Press.

Gates Foundation. (2014). *Using mobile data for development*. Retrieved from www.gatesfoundation.org

Gloss, A. E., Thompson, L. F., & McCallum, S. (2015). Putting human capabilities to work: The role of occupational information in global development. In I. McWha, D. C. Maynard, & M. O. Berry (Eds.), *Humanitarian work psychology and the global development agenda: Case studies and interventions*. London: Psychology Press.

Gomez, R., Baron, L. F., & Fiore-Silvast, B. (2012). The changing field of ICTD: Content analysis of research published in selected journal and conferences, 2000-2012. Paper presented at ICTD 2012: The International Conference on Information and Communication Technologies and Development, Atlanta, United States.

Griffith, E. (2014, 1 August). Can you find founders before they know they are founders? *Fortune*. Retrieved from www.fortune.com

Griffiths, R., Chingos, M., Mulhern, C., & Spies, R. (2014). *Interactive online learning on campus: Testing MOOCs and other platforms in hybrid formats in the University System of Maryland*. Retrieved from ww.sr.ithaka.org/

G20. *Seoul Multi-Year Plan of Action*. Retrieved from http://www.g20.utoronto.ca/2010/g20seoul-development.html

Hey, T., Tansley, S., & Tolle, K. (2009). *The fourth paradigm: Data-intensive scientific discovery*. Retrieved from http://research.microsoft.com

IBM. (2014). *Performance and capacity implications for big data* (Red Paper). Retrieved from ibm.com/redbooks

ILO (International Labour Organization). (2013). *Work of Work Report, 2013: Repairing the economic and social fabric*. Geneva: ILO.

ILO (International Labour Organization). (2014). *World of work report: Developing with jobs*. Retrieved from www.ilo.org

ITU (International Telecommunications Union). (2013). *Big data: Big today, normal tomorrow*. Retrieved from www.itu.int

ITU (International Telecommunications Union). (2014). *Measuring the information society report*. Retrieved from www.itu.int

Klinger, B., Khwaja, A. I., & Carpio, C. (2013). *Enterprising psychometrics and poverty reduction*. doi: 10.1007/978-1-4614-7227-8

Kolowich, S. (2013, 16 December). edX drops plans to connect MOOC students with employers. *The Chronicle of Higher Education*. Retrieved from www.chronicle.com

Lefkowitz, J. (2008). To prosper, organizational psychology should… expand its values to match the quality of its ethics. *Journal of Organizational Behavior*, *29*, 439-453.

Lim, B. C., & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior, 27,* 403-418.

Locke, E. A., & Latham, G. P. (Eds.). (2013). *New developments in goal setting and task performance*. New York, NY, United States: Routledge.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, *90*(10), 59-68.

McKinsey & Company (2011). *Big data: The next frontier for innovation, competition, and productivity*. Retrieved from http://www.mckinsey.com/

Meyer, J. P., Becker, T. E., & Vandenberghe, C. (2004). Employee commitment and motivation: A conceptual analysis and integrative model. *Journal of Applied Psychology*, *89*, 991–1007.

Molapo, M., & Marsden, G. (2013). *Software support for creating digital health training materials in the field*. Paper presented at ICTD 2013: The International Conference on Information and Communication Technologies and Development, Cape Town, South Africa.

Müller-Riedlhuber, H., & Ziegler, P. (2012). DISCO II – Prospects and challenges of a multilingual skills terminology. Retrieved from www.disco-tools.eu

OECD (Organization for Economic Co-Operation and Development). (2009) *21st century skills and competences*. Retrieved from www.oecd.org

OECD (Organization for Economic Co-Operation and Development). (2012). *Better skills, better jobs, better lives: A strategic approach to skills policies*. doi:10.1787/9789264177338-en

OECD (Organization for Economic Co-Operation and Development). (2013a). *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. Retrieved from http://skills.oecd.org/

OECD (Organization for Economic Co-Operation and Development). (2013b). *Exploring data-driven innovation as a new source of growth: Mapping the policy issues raised by "big data*." Retrieved from www.oecd.org

OECD (Organization for Economic Co-Operation and Development). (2013c). *New data for understanding the human condition*. Retrieved from www.oecd.org

OECD (Organization for Economic Co-Operation and Development), World Bank, ETF (European Training Foundation), ILO (International Labour Organization), & UNESCO (United Nations Educational, Scientific, and Cultural Organization). (2013). *Indicators of skills for employment productivity: A conceptual framework and approach for low-income countries* (A report for the Human Resource Development Pillar of the G20 Multi-Year Action Plan). Retrieved from http://www.oecd.org/

Pellegrino, J. W., & Hilton, M. L. (2012). Education for life and work: Developing transferable knowledge and skills in the 21st century. Retrieved from www.nap.edu

Reda, A., Shah, S., Tiwari, M., Lillie, A., & Noble, B. (2012). *Social network in developing regions*. Paper presented at ICTD 2012: The International Conference on Information and Communication Technologies and Development, Atlanta, United States.

Rose, I. (2014, 24, October) *Over 40? Data says you could be the perfect entrepreneur*. Retrieved from www.bbc.co.uk

Salas, E., & Cannon-Bowers, J.A. (2001). The Science of Training: A Decade of Progress. *Annual Review of Psychology*, *52*, 471–499.

Salas, E., Tannenbaum, S. I., Kraiger, K,. & Smith-Jentsch, K. A. (2012). The science of training and development in organizations: What matters in practice. *Psychological Sciences in the Public Interest*, *13*(2), 74-101.

SAS. (2011). *Using social media and online conversations to add depth to unemployment statistics*. Retrieved from www.unglobalpulse.org

Searle Center. (2014). *Searle Center Leadership Roundtable on Talent Analytics and Workforce Science (meeting agenda)*. Retrieved from www.law.northwestern.edu

Seibert, S. E., Wang, G., & Courtright, S. H. (2011). Antecedents and Consequences of Psychological and Team Empowerment in Organizations: A Meta-Analytic Review. *Journal of Applied Psychology*, *96*, 981–1003.

Sen, A. (1999). *Development as freedom*. New York: Knopf.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2004). *Experimental and quasi-experimental design for generalized causal inference*. Boston, MA, USA: Houghton Mifflin.

Spreitzer, G. (2008). Taking stock: A review of more than twenty years of research on empowerment at work. In J. Barling, & C. Cooper (Eds.), *The SAGE handbook of organizational behavior: Volume I - micro approaches* (pp. 54-73). doi:dx.doi.org/10.4135/9781849200448.n4

Suddaby, R. (2014). Why theory? *Academy of Management Review*, 39(4), 407-411.

TechAmerica Foundation. (2014). *Demystifying big data: A practical guide to transforming the business of government*. Retrieved from www.techamerica.org

Tippins, N. T., & Hilton, M. L. (2010). *A database for a changing economy: Review of the Occupational Information Network* (O*NET). Retrieved from http://www.nap.edu/

UN (United Nations). (2013). *Big data and modernization of statistical systems* (Economic and Social Council E/CN.3/2014/11). Retrieved from www.unstats.un.org

UN (United Nations). (2014). *The Millennium Development Goals Report 2014*. Retrieved from http://www.un.org/

UN Global Pulse. (2012). *Big data for development: Challenges and opportunities*. Retrieved from http://unglobalpulse.org

UN Stats (United Nations Statistics Division). (2013). *Big data for policy, development, and official statistics* (Friday Seminar on emerging issues). www.unstats.un.org

UNDP (United Nations Development Programme). (2014). *The role of the private sector in inclusive development: Barriers and opportunities at the base of the pyramid* (A report by the Istanbul International Center for Private Sector in Development). Retrieved from http://www.undp.org/

UNESCO (United Nations Educational, Scientific, and Cultural Organization). (2007). *ICT in education programme*. Retrieved from www.unescobkk.org

UNESCO (United Nations Educational, Scientific, and Cultural Organization). (2012). *Shanghai Consensus: Transforming technical and vocational education and training - building skills for work and life* (Recommendations of the Third International Congress on Technical and Vocational Education and Training – Shanghai, People's Republic of China). Paris, France: UNESCO.

UNESCO (United Nations Educational, Scientific, and Cultural Organization). (2014). *Institute for Lifelong Learning Medium-Term Strategy* (2014-2021). Retrieved from www.unesco.org/uil

UNEVOC (International Centre for Technical and Vocational Education and Training). (2013a). *ICTs for TVET: Report of the UNESCO-UNEVOC online conference*. Retrieved from www.unevoc.unesco.org

UNEVOC (International Centre for Technical and Vocational Education and Training). (2013b). *Revisiting global trends in TVET: Reflections on theory and practice*. Retrieved from www.unevoc.unesco.org

United States Department of Labor. (2014). What works in job training: A synthesis of the evidence. Retrieved from

Unger, J. M., Rauch, A. Frese, M., & Rosenbusch, N. (2011). Human capital and entrepreneurial success: A meta-analytical review. *Journal of Business Venturing*, *26*, 341-358.

Van Deursen, A. J. A. M., Courtois, C., & Van Dijk, J. A. G. M. (2014). Internet skills, sources of support, and benefits from Internet use. *Human-Computer Interaction*. doi: 10447318.2013.858458

Van Dijk, J. A. G. M. (2005). *The deepening divide: Inequality in the information society*. Thousand Oaks, CA, USA: Sage Publications, Inc.

Venkatesh, V., Morris, M. G., Davis, G. B., Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425-478.

Wainer, H. (2000). Introduction and history. In H. Wainer (Ed.), *Computerized adaptive testing: A primer* (2nd Ed.). New York, NY, USA: Routledge.

Welter, F. (2011). Contextualizing entrepreneurship: Conceptual challenges and ways forward. *Entrepreneurship Training and Practice*. doi: 10.1111/j.1540-6520.2010.00427.x

West, D. M. (2012). *How mobile technology is driving entrepreneurship*. Retrieved from www.brookings.org

White House. (2014). *Big data: Seizing opportunities, preserving values*. Retrieved from www.whitehouse.gov

World Bank. (2012a). *Jobs*. Washington, DC: World Bank. doi:10.1596/978-0-8213-9575-2

World Bank. (2012b). *The Right Skills for the Job?: Rethinking Training Policies for Workers*. Washington, DC: Author. doi: 10.1596/978-0-8213-8714-6

World Bank. (2012c). *Maximizing mobile.* Retrieved from www.worldbank.org

World Bank. (2014a). *STEP Skills Measurement: Snapshot 2014*. Retrieved from www.worldbank.org

World Bank. (2014b). *Big data in action for development*. Retrieved from data.worldbank.org

World Economic Forum. (2013). *The Human Capital Report*. Retrieved from www3.weforum.org

World Economic Forum. (2014). *The global information technology report: Rewards and risks of big data*. Retrieved from www3.weforum.org

Zahra, S. A., Korri, J. S., & Yu, J. J. (2005). Cognition and international entrepreneurship: Implications for research on international opportunity recognition and exploitation. *International Business Review*, *14*(2), 129-146

Zheleva, M., Schmitt, P., Vigil, M., & Belding, E. (2013). *Bringing visibility to rural users in Cote d'Ivore*. Paper presented at ICTD 2013: The International Conference on Information and Communication Technologies and Development, Cape Town, South Africa.