A DEMOCRACY OF UNEQUALS:
SOCIAL DIFFERENTIATION, PARTICIPATION INEQUALITY
AND THE COLLABORATIVE IDEAL ONLINE

Robert Bruno, Ph.D.
Purdue University
November 1, 2010
As the most social of creatures, humans are shaped more by their interactions with people than any other factor. We learn not only of the world, its constituent elements, and our place in it during this social process, but also what sets us apart. Thus, human interaction, and the development of knowledge associated with it, is both a process of assimilation and differentiation. It follows that the most powerful kind of learning is not a solitary, selfish endeavor but one that takes place within the context of groups made up of myriad individuals with a shared objective. However, this path to enlightenment is not one defined only by equanimity but often by constant negotiation between those assuming unequal positions within a structurally hierarchical landscape. To be sure, human behavior is extremely complex and oftentimes surprising. Yet, seminal communication research in the mid-19th century showed it in fact to be predictable, and quantifiably so. Social interaction reveals itself to be unequal in many ways, especially within groups working on a task. It has long been known that groups functionally differentiate (Bales, 1951; 1953; Stephan and Mishler, 1952; Slater and Bales, 1955) in terms of structure, status, roles and the amount of participation their members contribute to collective activities and recently this has also been shown to occur within technologically-mediated spaces, perhaps more dramatically (Shirky, 2000; 2004; 2008; Kumar, Novak, and Tompkins, 2006; Kuk, 2006; Kittur, Chi, Pendleton, Suh, and Mytkowycz, 2006; Ortega, Gonzalez, and Robles, 2008; Bryant, Forte, and Bruckman, 2005; Sunstein, 2006; Raymond, 2001). What transpires in this naturally-occurring scenario is that individuals, through their uneven social actions, reactions and the roles they assume, become quantitatively differentiated.

It is also well known that within a collaborative task group the most powerful kind of learning is occurring in tandem with these social interactions (Slavin, 1996; Johnson and Johnson, 1998), a gestalt process whose outcome is more than the sum of its individual parts.
The notions of social interaction and social learning are actually not discrete, especially considering that much foundational learning theory is informed by conceptualizations of communication (Smith and Ragan, 2005; Shannon, 1948). Accordingly, the novel idea proposed in this paper is that the most powerful learning outcomes of all take place in concert with optimal levels of social differentiation.

**Theoretical Framework.** This is rooted in the notions of social learning and social differentiation. Social learning theory draws from constructivism and is underpinned by decades-long developmental psychology and collaborative learning research which provides the evidence for powerful learning effects resulting from group collaboration. Social differentiation and structuring simply occurs when individuals contribute varying inputs to group communication processes. However, despite indisputable proof that this consistently occurs, its existence alone is not what is truly notable – but the undeniable fact that these structured arrangements not only work, but work better, than their unstructured counterparts. Research is just starting to show that uneven internal processes are more efficient and lead to better outcomes, whether it be more accurate Wikipedia articles, higher quality open source software code (Kuk, 2006; Kittur et al., 2005) or more powerful learning.

Communication and learning are inextricably linked. Notions of constructivist and developmental learning are based on the communication process and individuals’ interactions with others. This is also true of the more formal educational paradigms of collaborative and cooperative learning as well as other learning theories which explore leader-follower effects and communication-participation levels on group performance. Discussing participation dynamics with respect to their learning effects allows for an easy transition into a deeper consideration of the relationship between social learning and social differentiation/participation inequality. This
paper makes the case for an intrinsically potent relationship between participation levels and learning – relative participation levels in small and large task and/or collaborative groups that are naturally occurring, structured, unequal, efficient and predictable.

The concept of communication-participation inequalities, the processes that generate them, and their exemplifications in human interaction have been recognized in myriad shapes and been identified through numerous descriptors, including “asymmetry,” “imbalance,” “differentiation,” “structuration,” “self-organization,” “power law distributions,” etc. Probably the most well-known of these is the Pareto Principle, also known as the “80/20” rule. Originally applied in economic contexts as the regularly-occurring unequal distribution of wealth among populaces, the resultant ratio appears to apply to innumerable kinds of social interaction. In Here Comes Everybody (2008), Shirky noted that these principles and outcomes especially apply to social interactions online: it is a structure that is native to social media across a wide array of interactions and is consistently a positive system attribute (pp. 124-5).

Shirky also pointed out that “most large social experiments are engines for harnessing inequality rather than limiting it” (p. 128). He does not view inequality as a bad idea in itself, but ultimately the best, most desirable dynamic to achieve. For if not promoting perfect equality, online environments certainly do foster another hallmark of democratic systems: a great degree of individual freedom – thereby allowing individuals of all types and talents wildly varying levels of participation, contribution, and effort, which ultimately appear to lead to the best outcomes. Tellingly, especially for the discussion at hand, Shirky noted that “power law distributions tend to describe systems of interacting elements, rather than just collections of variable elements” (2008, p. 128)…(and that the) imbalance appears in an astonishing number of places in…social systems” (p. 126).
Ortega, Gonzalez, and Robles, in “On the Inequality of Contributions to Wikipedia” (2008), conducted a quantitative analysis of contributions to the large online encyclopedia Wikipedia which showed that the vast number of its articles are composed and edited by a small number of participants. It was demonstrated that this distribution resembles the Pareto Principle (the 80/20 rule), power laws and the other notions of inequality discussed above. Furthermore, in “The Structure of Social Collaboration on Wikipedia,” Matei, Braun, and Petrarche (2009) confirmed this phenomenon of unevenness and also that it becomes “hardened” over the course of time and number of edits per article. In truth, there has been a large amount of similar research discovering social imbalances in which Wikipedia is the favored analytical territory (Sunstein, 2007; Shirky, 2008; 2004; Fuchs and Gasse, 2008; Bryant et al., 2005; Voss, 2005; Kittur et al., 2005; Schwartz, 2006), but the phenomenon is not just limited to this particular online space. Many socially-driven virtual properties display unevenness as well, including blogs, online interest groups, sites that utilize user-generated content, phishing Web sites, mailing lists, online communities, social networks, peer review sites, open source software development, bulletin boards, online games and many, many more (Shirky, 2004; Raymond, 2001; Moore and Clayton, 2008; Kuk, 2006; Madey et al., 2002; Mockus et al., 2002; Kumar et al., 2006; Correia et al., 2006; Huberman, 2001). As described by Shirky, this skewness seems to be innate, and the norm, in online social interaction, not an anomaly.

Yet these uneven dynamics which exist in large-scale online systems were identified decades earlier in face-to-face small group interaction, most notably by Bales in “The equilibrium problem in small groups” (1953), among other relevant works, and Stephan and Mishler in “The Distribution of Participation in Small Groups: an Exponential Approximation” (1952). Bales conducted several small group discussion/problem solving experiments (content
analyses) with his colleagues at Harvard University. He noted: “The balance of action with reaction is one of the equilibrium problems of the system…a problem of a certain balance in the way in which these activities are distributed between separate members” (p. 117). Bales found this consistently across varied groups of different sizes, along with recurrent patterns and regularities. He concluded this particular research effort by saying, “It may very well be that the expectation of ‘equality’ which is so often present in groups of our culture…(in) practice is never found” (p. 131).

Stephan and Mishler (1952) built on Bales’ findings and premise by attaching the respective quantitative levels derived through experimentation to individual relative participation in differently sized groups. Notably, they specifically analyzed the process of group collaboration in a learning situation and counted words, sentences and statements as the operational units of participation. They confirmed Bales’ findings of inequality while further modeling the patterns within group interaction.

A system perspective. These particular patterns continue to manifest as this discussion moves to larger scales of analysis. In “Groups as social systems” (1975), Berkowitz summed up much of the previous discussion: “In one way or another, the people in the group become differentiated from each other…When social psychologists discuss these stabilized relationships, they talk about the group structure, the relatively persistent pattern of relationships among the differentiated parts of the groups” (p. 464). Yet, it is Berkowitz’ choice of title for this chapter which he explained is about groups as “functioning units” that may be most telling. Social phenomena began to be examined and explained through system constructs drawn from technology and the physical sciences by mid-20th century, right around the time Bales (who regarded groups as “microscopic social systems,” 1950), Stephan and Mishler (1952), and others
were conducting their research on small groups. Although sometimes separated within the
discussion for the sake of order and simplicity, an operating assumption of this paper is that these
same principles apply equally to small groups and larger systems.

Frey further articulated the connection between small groups and systems in *The Systems
Metaphor in Group Communication* (1999):

One of the most dominant, and probably the most taken-for-granted,
constructs applied to the study of small groups is system-ness…a core
assumption of systems theorizing is that communication is the observable
phenomenon binding together constituent components of systemic
entities…The concept of information, developed originally in
mathematical statistics for applications in the study of electrical
engineering, quickly found its way into social-scientific theories and
research (pp. 71-73).

As other researchers in this realm have concluded, ultimately what is important are not
necessarily the underlying mechanisms that drive such differentiation – i.e., the qualitative nature
– although these will be examined at some length below, but instead the recurrent quantitative
character of these social dynamics in small and large groups, especially their unevenness.

**Role differentiation and rates of participation.** Role differentiation provides the process
through which the evidence of naturally-occurring social inequalities manifests – consistent rates
of relative participation within any and all groups. Much of Bales’ early theoretical positions
came about because of the preceding era’s long-established assumption of the existence of
balance in systems. However, he found small group systems are *not* balanced. He explored this
conundrum in more depth in “The equilibrium problem in small groups” (1953). Previously,
Bales (1950) had tracked both positive and negative feedback (or reactions) to group member actions and noted that positive feedback greatly outweighed negative. Consequently, those who were more active tended to receive the most positive feedback and were thus recognized as “leader” types. In this self-reinforcing process, clearly there is a correlation between activity and status – from top to bottom, status rank order equates neatly with diminishing proportions of contributions – and therefore a lack of balance. Obviously, the most prominent signs of this are the stark divides that occur in participation and status themselves.

*The social Internet and self-structuring systems.* While many system theories may help make plausible sense of large-scale social structuring and inequality, the empirical examples below provide the true evidence of how these processes have taken real shape online. Structuring processes and imbalances materialize in numerous large-scale ways on the Internet, nearly all involving collaborative interaction of some manner. These most prominently include social applications online, Wikipedia and wikis, and the open source software movement (OSS). From this perspective it can be discerned how, given that technology acts as a catalyst for these social structuring processes, the Internet itself is a self-organizing system wherein mediated communication enables interaction.

In “The Internet as a Self-Organizing Socio-Technological System,” Fuchs (2005, p. 1) argued that because “the technical structure is medium and outcome of human agency” the Internet itself is a self-organizing system, fulfilling Giddens’ required reciprocal framework for structuration theory. According to Fuchs, “Human actors permanently re-create this global knowledge storage mechanism by producing new informational content, communicating in the system, and consuming existing informational content in the system” (2005, p. 9).
Shirky has noted that “this (self-organizing/structural) pattern (of uneven distributions) is general to social media – imbalance is the same shape across a huge number of different kinds of behaviors” (2008, p. 124), and the behaviors of which he speaks are specifically communicative in nature. Oftentimes communication more fully constitutes participation and contribution in online settings as opposed to co-located and/or face-to-face scenarios, specifically textual communication. To state the obvious, non-communication-based interaction is nearly non-existent online. Shirky (2004; 2008) has speculated that inequalities exist online due to freedom of choice and a greater diversity of roles to assume, which ubiquity and access to technology make possible. “The freedom driving (activity online) removes the technological obstacles to participation. Given that everyone now has the tools to contribute equally, you might expect a huge increase in equality of participation. You’d be wrong” (2008, p. 123). Although Shirky’s contention that “everyone now has the tools” is certainly debatable, his observations about the actual collective that does participate are keen: “Large social systems cannot be understood as a simple aggregation of the behavior of some nonexistent ‘average’ user” (2008, p. 125). This is all the more true because in frequency distributions that describe this kind of social behavior, the vast majority of participants contribute less than the mean. Simply, a few participants contribute much more than everybody else.

**Wikipedia, Wikis and Open Source.** Wikipedia is the most prominent example of massive knowledge creation on the Internet and that it is a working system based on the “open source” philosophy itself, which might work *better* than its print counterparts developed through traditional paradigms, is significant. Analyzing the immense user-generated online encyclopedia allows for a valuable case study of one of the Internet’s true early-stage interactive wonders as well as provides a legitimate basis to make inferences about wikis as social and learning
mechanisms, in general. Wikis epitomize the most simple, open elements of the Internet – they are easy-to-use, and enable anyone to read from, write to, add, delete or edit any of the content on their pages. They are engines of the user-generated Web and offer very few barriers to participation and collaboration between different users.

Wikipedia has proven fertile ground for research and, not surprisingly, wiki platforms are gaining traction as actual research environments themselves. With the breadth and depth of its online encyclopedia and huge user base, Wikipedia is a prime example of massive collaboration. Although Wikipedia does assign pre-determined roles for a select few (editors, administrators, etc.), and that certain rules and protocols (basic contribution rules, a Neutral Point Of View, etc.) exist, these guidelines are more often viewed as accepted norms of “Wikiquette” to abide and not hard-and-fast laws that users must obey. Wikipedia still exemplifies a primarily implicit system of low regulation and absence of close management, staying true to its ethos of openness. Therefore, it remains a valuable source for empirical evidence of naturally-occurring, large-scale structuring processes, inequality and self-organization – and, as will be demonstrated, certain types of outcomes as well.

In a theoretical essay, Matei and Bruno (2008) examined the editing and “reverting” (returning the article text to a previous version) process used in Wikipedia and proposed that it is shaped by conflict. However, although conflict defines much online interaction it also leads to resolution and stability, and the overall evolution of the system: “A study of collaboration practices in a free software community revealed that conflict is at the heart of the open-source collaborative process, where it plays an equally important role in structuring and renegotiating rules and principles” (2008, p. 20). The researchers also speculated that “de facto” hierarchies
arise from the online encyclopedia’s loose protocols and an agreed-upon understanding of acceptable behavior within the online community.

At heart, Wikipedia is a knowledge source that allows its own users to participate in the very production of that knowledge. Although in general eschewing the assistance of experts, the depth, breadth and accuracy of its content is still the ultimate measure of its value. But it is this lack of a strictly enforced structure (closer to looser, “naturally-occurring” protocols that arise from interaction), combined with the power of large numbers that gives Wikipedia its legitimacy, perhaps even outperforming traditional knowledge sources and their more mainstream models of production. These large numbers and their benefits are enabled because “(it is an) accepted reality that online collaborative environments have lowered the barriers to entry, both technical and social, (which) will encourage communal editing either of content or of policies, via continuous collective vigilance and self-monitoring (Matei and Bruno, 2008, p. 19).

Shirky (2008) has conjectured that large-scale inequalities and structuring in fact do exist due to similar dynamics as those on the micro-scale, in small groups. He has called this “flexibility of role,” a freedom that is readily available to those choosing to participate in large online initiatives like Wikipedia. For example, one can easily move between being both a consumer and producer of content, a reader, editor or, if designated so, an administrator of some kind. Again, these distinctions clearly mirror those of Bales (1950), in that most active leaders will emerge (administrators), but that the majority of participants will contribute through lesser, varied and peripheral task activities.

Shirky has made the claim that collaborative systems like Wikipedia work because of their inherent inequalities, not in spite of them. According to him, there should be no call to even out the contributions among participants because that would destroy the system. “No effort is
made to even out…contributions. The spontaneous division of labor driving Wikipedia wouldn’t be possible if there were concern for reducing inequality” (2008, p. 125).

**Open Source Software.** If Wikipedia is simply one of the most prominent parts of the rising new media landscape – a domain where self-organizing and natural structuring processes have come to embody the recognizable shapes of virtual, working social systems – the open source software realm can be regarded as the birthplace providing the “primordial soup” from which these processes of genesis animate and take form. Like Wikipedia, an OSS model itself, the open source software development movement provides a rich source and landscape for research into large systems, their inherently unequal structuring/self-organizing processes, how these systems work and how they may in fact work better than their more traditionally-structured counterparts. Sunstein (2006) stated, “Much of the excitement of the open source movement stems from a simple fact: If we understand why open source software works well, we might be able to use the open source idea in many other contexts” (p. 18). The most identifiable illustration of why this question is relevant is the wild success of Linux software, the open source model, rivaling the dominant, monolithic, and closed model for production, Microsoft Corporation, for market share.

Open source software development involves large groups of disparate, geographically-dispersed individuals working together usually with no central, guiding authority. However, lacking clear goals and direct management seems to be advantageous in these contexts because out of these differences a “coherent and stable” system magically appears, according to Sunstein (2006, p. 172). He compellingly offered an explanation of open source groups that are equivalent to their best members or groups that become better than their best members through processes of learning and synergy: “The give and take of group discussion might sift information and
perspectives in a way that leads the group to a...solution...in which the whole is actually more than the sum of the parts...a process of synergy or learning, spurring creativity and producing an outcome that is far better than a mere aggregation of existing knowledge” (2006, p. 55). Much of these dynamic processes take place due to statuses, identities and the individual contributions of the participants involved, who are oftentimes divided into roles within a core of active users (leaders) and periphery of sometime (ordinary) users – in a system lightly bounded by both formal and informal rules (Sunstein, 2006).

In *The Cathedral and the Bazaar* (2001), Raymond surmised that the best work is done by software developers who can tap into the best of what the collective offers: “The really great hacks come from harnessing the attention and brainpower of entire communities...the developer who knows how to create an open, evolutionary context in which feedback...and other improvements come from hundreds of people” (2001, pp. 50-1). It is difficult to avoid recognizing how the discussion of these OSS systems echoes the earlier literature review of constructivism and collaborative learning. Raymond also suggested the potential usefulness of defined goals, a level of monitoring and motivational activity, and the outlines of a functioning structure to best marshal available resources.

Participation inequality (PI) and outcomes are formally investigated by Singh et al. (2007), who found that PI is common to open source development and correlated in with project efficiency: “It is observed that projects with a clearly identified core group of developers conducting most of the coding are more efficient for project progress. We also find that a highly participative peripheral group in communication positively affects project progress” (p. 1). However, of all the literature reviewed on group interaction, unequal and peripheral participation, the Internet and OSS structuring, none go deeper and reveal these processes better
than Kuk (2006), who conducted research and conjectured on all of the above, but also questioned why and to what end (he made hypotheses about “optimal” levels). Kuk reviewed much of the literature revealing extreme cases of participation inequality in code development and its adjunct activities, e.g., “the most prolific 1% contributed 20% of the postings, and the top 20% contributed 61% of the messages…in a study of Gnome (an open source project), (it was) found that 17% (52 of 301 programmers) contributed 80% of the code” (2006, p. 1031). Kuk cited Bales (1950; 1953; Bales and Slater, 1955) directly as the theoretical forbear for explaining this modern phenomenon: “Participation inequality is a well-known problem in…groups…It underlines a social and psychological phenomenon that distribution of participation is not random but hierarchical, with a few highly verbose members dominating the discussion even when the conditions encourage equal participation” (Kuk 2000; 2006).

Prior to conducting his research, Kuk (2006) attempted to explain the potential reasons behind inequalities through notions of the production of goods and collective efforts: “Recent empirical and theoretical research in collective action suggests…that participation inequality is essential and serves as a productive lever by…mobilizing personal and network resources” (p. 1032). He noted that individuals do not connect with others randomly but self-select into collective arrangements with those most resourceful and the best chance for success. Kuk aligned these activities to the notions of communities of practice, active participation and legitimate peripheral participation (Lave and Wenger, 1991), as well as the ideas that decentralized management leads to greater efficiency, structures of operation emerge, and a holistic process encourages others to participate: “The developers who remain cognizant of the latest developments and who post the most interesting and innovative ideas are most likely to captivate
attention and persuade others that their reciprocations will further enhance collective competence” (2006, 1033).

**The Argument for Inequality.** Given all these phenomena, the evidence of recurrent inequalities, and the potential for its significant effects, how to explain it all at a higher level of abstraction? Interestingly, despite not conducting true scientific analyses himself, it may be Shirky (2008; 2004; 2000) who has made the most eloquent arguments for the positive benefits of participation inequalities in online environments. As noted previously, Shirky pointed out that “imbalance drives large social systems rather than damaging them,” (that) “social experiments are engines for harnessing inequality rather than limiting it” (2008, pp. 124-128), and that efforts setting out to reverse participation inequalities are “tantamount to killing the system” (Correia, 2006, p. 13). Shirky contended that such overt action is not only wrongheaded but damaging: “Because it arises naturally, changing this distribution (of participation) would mean…the application of force. Reversing the…system would mean destroying the village in order to save it” (2004, p. 50).

In “In Praise of Freeloaders” (2004), Shirky did not point out the benefits of the most active or leaders, but actually those on the other end of the participation spectrum, the least or non-active members of a group – free riders, individuals who benefit from the group activity but appear to not sufficiently contribute to the good of the whole in what some scholars regard as a tragedy of the commons scenario. Shirky contended (and Kuk concurred) that although their contributions may not be overt, these individuals are in some manner essential to the functioning of the system.

Inevitably, however, the idea of inequality leads to questions of fairness, to which Shirky has responded, “asking whether there is inequality…is the wrong question, since the answer will
always be yes. The question to ask is, ‘Is the inequality fair?’’ (2004, p. 50). According to Shirky, inequalities are not only fair, but the best state of affairs possible. From an operative standpoint, free riders may not be totally inactive but instead involved in ancillary activities which are indirectly linked to the task at hand (e.g., legitimate peripheral participation). To denigrate these non-actors, or worse to attempt to correct or eliminate them altogether, goes against the holistic concept of what a system is and how it functions. The least active organs are as important as the most active organs in keeping the “body” of any system alive, well and functioning properly. With regard to those at the other end of the scale, attempts to force equity or stifle the most active often fail and are counterproductive in and of themselves. The most active participants, which include leaders and top performers, clearly raise the performance of the entire group.

**Shannon’s entropy index: a superior measure of online communication.** Shannon’s entropy index, as incorporated into the Visible Effort Wiki (see http://veffort.us), is believed to be the best measure available for participation inequality in online communicative interaction. The entropy index is drawn directly from information and communication theory (how much information or meaning is carried within an instance of communication) and was derived from Shannon’s work on technological systems. Entropy is equally adept at measurement of small and large groups (allowing for system-level analysis), another distinguishing characteristic from other measures and directly applicable to the research conducted here which examined small group interaction. As designed for the Visible Effort Wiki, the entropy index, which is in essence a mathematical ratio, allows for both net and gross calculations of overall relative participation (operationally, textual characters contributed to a collaborative space) levels as well as calculates and stores both individual and group-level information. Although the index can be broadly and
liberally applied (for comparison purposes between online systems), it can also be valued as a simple, central and stand-alone statistical measure – not unlike the median or mean – which informs the researcher of something unique about a distribution of values. Most important is that, from a theoretical and operational standpoint, meaningful messages and their symbol distribution use words and letters in an uneven manner and the entropy index is a measure of this disorganization and its respective meaning (Matei et al., 2009). Or, put more simply, communication is in itself inherently imbalanced, which makes the entropy index its most suitable gauge.

**Optimal levels.** This paper offers the hypothesis that there is an optimal level of participation inequality in relation to a performance outcome. Social interactions lead to social differentiation and inequality, and that inequality takes a general shape along a continuum. Low differentiation and high differentiation in interaction within groups can both lead to problems, and the idea of optimality is of an ideal level operation in between these two states. A great deal of direction is taken from Kuk’s (2006) and Matei et al.’s (2006) leads. To this author the idea of an optimal level of online participation was first identified in “Collaboration and Communication: A Social Entropy Approach” (Matei et al., 2006, p. 18):

> As groups gel and functional hierarchies are formed, entropy will not be maximized, but rather reduced or maintained at an optimal level. From this perspective, one of the most important theoretical contributions of research conducted from a social entropy perspective would be to find what entropy levels are optimal for any given online communicative and collaborative system. Furthermore, which entropy level corresponds to what level of functional differentiation and hierarchical organization?
While it may be one matter to find and establish recurrent imbalances in small and large group collaboration which attain, maintain and seemingly thrive in these stable modes of functioning, it is still another to determine what might exactly be an “optimal” state of operation. First of all, by what criteria and measure is this optimal state to be evaluated? Usually, it is by gauging some kind of outcome that is expected or, in experimental terms, a dependent variable. Fortunately, in most collaborative efforts, there is a clear goal with a specific end product in mind. In the case of Kuk (2006), the amount of knowledge sharing is the dependent variable quantified resulting from the communicative interactions between open source software developers on teams of varying levels of participation inequality.

The dependent variable under examination in this paper is learning gain. It is believed that, as in Kuk (2006), optimal participation inequality levels can be hypothesized beyond which performance levels (learning gain) start to go down. It is ventured that an optimal entropy level exists with accompanying relative participation proportions that mirror those found by Bales (1950; 1953) and other early researchers, as well as Kuk (2006) and recent researchers – the presence of one or a few participants who are significantly more, or less, active than the others.

Although academic papers detailing participation inequality or variables affecting group collaborative performance are abundant, a comprehensive, an in-depth literature review revealed very little scholarly research bridging the two. Of that, which includes much from the prolific social commentator Clay Shirky (2000; 2004; 2008), only Kuk (2006) provides empirical evidence for a curvilinear relationship between participation levels and an outcome derived from scientific analysis.

When a relative participation level is highly correlated with an ideal outcome, then it has attained that level itself – the most powerful learning is associated with this optimal participation
level. Operationally, the optimal level is defined in the research undertaken here as when participation inequality/evenness levels and learning gain, which to a certain point have generally maintained a linear relationship, reach an inflection point with learning gain (the dependent variable) beginning to decline. This peak, or range, would constitute an optimal participation inequality level for highest learning gains.

Logically, this optimal level of interaction would stabilize and not diverge significantly from what is attained through naturally-occurring structuring processes. On the other hand, it would be unreasonable to assume that this equilibrium point in itself would always be an ideal level of operation. Any experienced manager or teacher knows that employees and students charged with delivering a product as a task group will need the right combination of guidance and encouragement to produce the best results. However, an assumption of the model proposed here is that the positive benefits of allowing for self-organization may outweigh heavy-handed intervention. As it specifically relates to the research undertaken (see Method), the individuals in a small task group assumed different levels of text contributions, with some participating much more and others much less. To stifle the most active, or high performers, is a misguided form of regulation as the actions of these individuals raise the performance of the entire group; likewise, the less active may be contributing something intrinsically valuable themselves, if less overtly.

The entropy index gives a system-level measure for group participation evenness/inequality while the Visible Effort Wiki also displays individual relative participation levels. It was predicted that, as found by both early researchers (Bales, etc.) and recent ones (Kuk, etc.), that there would be one or more most active participators and some who are much less active as well. In the quest for that “holy grail” of optimal levels, Matei et al.’s (2006) line of thought was taken – that the groups will have naturally attained optimal relative participation interactions.
Despite this tantalizing prospect, very little research has delved into optimal states online and fewer still have taken the approach of correlating performance outcomes with participation levels (one of the obvious reasons for this dearth has been until now the perplexing question of how to measure online participation?) The exception is Kuk (2006). Not only does Kuk connect many of the conceptual dots from Bales, other early researchers and some more recently, he affixes it to a performance outcome in the most modern of collective endeavors, open source software development.

Kuk (2006) conducted a content analysis on a sample of 867 postings to an open source software discussion list. From the data, he was able to glean the group’s level of knowledge sharing, a performance variable closely associated with the development and quality of computer code, and the level of overall participation inequality. In analyzing the data, Kuk found a strong curvilinear relationship between levels of participation inequality and the level of knowledge sharing between code developers. Beyond a certain level of increasing inequality, the level of knowledge sharing starts to go down – basically revealing an optimal level of participation inequality for open source software development. Kuk advanced our understanding of the collective processes and underlying dynamics of OSS development (2006, p. 1031) by surmising that developers were more productive in smaller, better organized groups in which communication was networked to other peripheral developers most efficiently. He called this strategic interaction. The differences in relative levels of participation seemed to increase along with knowledge sharing via strategic interaction. At the same time, Kuk predicted that extreme concentrations of development would have the opposite effect, which is exactly what was found.

Model. It is assumed that social interaction and social learning are correlated. Thus, participation evenness in group collaboration and learning gain have a curvilinear relationship,
with evenness and learning rising in conjunction with each other, and learning eventually peaking and dropping while entropy continues to rise. There is, therefore, an optimal participation inequality range with regard to most powerful cognitive effects.

**Hypothesis.** Participation inequality/evenness, as measured by the entropy index, and learning gain will exhibit a curvilinear relationship, first rising, then declining (an inverted U).

**Method.** To test this hypothesis, an experiment was conducted using Visible Effort. Students were recruited and tasked to co-create online glossaries of terms as a group. The main goal was to empirically identify the degree to which collaborative evenness promoted learning or not. Learning was measured as acquisition of knowledge related to specific topics and facts addressed in the term glossaries.

The main presupposition behind this experiment was that learning outcomes improve as groups become differentiated. It was expected that as members started to contribute according to their individual level of knowledge, a specific social structure of learning would emerge. This structure offered each student a given role and comfort zone. Furthermore, it was expected that students would contribute in different ways, according to their needs, abilities, and motivations. The groups they participated in would be characterized by a specific level of collaborative differentiation and unevenness, which would go hand-in-hand with a specific level of learning effectiveness. It was further hypothesized that the relationship between learning outcomes, group structure and collaborative unevenness is curvilinear. When collaborative unevenness and its companion level of group structure reach the level where some of the group members constantly dominate the collaborative process or where too many members “free ride,” learning is disrupted. Group processes would be increasingly hindered by discussions and conflicts about optimal level of contribution, reward allocation, and equity. Collaboration would slow down or even cease.
However, the inverse would also be problematic. On the other extreme, collaboration could also become too even wherein top performers may not be allowed to stand above the others and consequently raise performance of the group whole.

170 students were recruited, randomly divided, and placed into 23 groups of between four and 11 individuals and each group was tasked to collaboratively define university-related terms over the course of a week. The level of group structuration for each collaborative team was measured using Visible Effort, which calculated the entropy value for the entire amount of contributions made to a specific topic and for the entire period of time dedicated to the study. All respondents were invited to take pretests and posttests, which measured learning. The main dependent variable of the tests was a learning gain index, which was calculated as the difference between posttest and pretest scores. Because the main independent variable is a group-level indicator (entropy), the dependent variable was aggregated at group level, the mean learning gain value being calculated for each separate group.

The main statistical analytic procedure employed was multiple regression. Learning gain was regressed on group entropy, controlling for a number of other factors, such as technical competency, which were selected from a number of possible factors that could influence learning gain. To test for the possibility of a curvilinear relationship, we also included in the equation the squared version of the entropy variable. The main finding of the research study is that there is indeed a curvilinear relationship between learning gain and group entropy, which was evidenced by the significant value of the overall model ($F(4,18) = 9.57$) and entropy squared term ($\beta = -6.23$, $R^2 = .14$) introduced in the equation to test for cuvilinearity. In more descriptive, bi-variate terms, as illustrated in Figure 1, learning gain increases up to an optimal entropy level of approximately 90. After this point, for each unit of entropy increase learning gain decreases. In addition to
entropy, learning gain was also influenced significantly by technical competency, but not raw participation level or knowledge of university-related information.

*Figure 1*: Bivariate curve estimation for group entropy and group learning gain.

**Discussion.** The main hypothesis of this study was supported – the analysis of data reveals there is a curvilinear relationship between relative participation in groups and learning gain. The claims made here are about groups and group interaction and the fact that the results are significant and positive maps well with the conceptual framework. In sum, the findings offer strong support for the relationship between participation levels in collaborative groups and learning.

This study determined that the quantitative character of collaboration matters, and that – as evidenced by the curve – differentiation matters as well. Some participants contributed more, some less, some in the middle range. Leaders emerged and free riders lurked. Using the wiki group with the highest learning gain as a prime example, it was noted that one user contributed 42% of the gross content to these definitions and two people contributed just 3 and 5% each. The remaining members in this group all fall into the middle range. Entropy for this group was 83.46,
not far from the mean of the frequency distribution. This without question supports the social
differentiation, inequality, leader/free rider and social learning theories put forth in this paper.
Analysis of this kind – positing operational benchmarks of relative participation – is a must have
in future research.

The theoretical implications of this research program are broad and large in scope. This
will become especially true as information communication technologies become more and more
insinuated in our lives and collaboration is not only encouraged but demanded as a way for
humans to overcome the most vexing challenges and issues advanced, deeply complex societies
are bound to face. On the most elemental level, quantified collaboration has been shown to relate
to outcomes. The applications of this idea are limited only by the imaginations of researchers and
the one undeniable conclusion arising from this study is that more research is warranted and
needed.

**Visible Effort as a practical tool.** This brings the discussion to a close highlighting a
primary element missing in the research conducted for this paper – active use of the Visible
Effort Wiki as a regulatory tool. The overt simplicity of the tool can cloak its ingenuity and
powerful utility – it is something that can be easily used “on-the-fly” by teachers and managers
to influence optimal collaboration levels and outcomes. These levels could be used as exemplars
in learning environments and beyond and no similar tool with regard to this kind of usage
presently exists. However, it is not proposed that the tool should be used in isolation, but in
concert with other tools and relevant collaborative and/or learning methods.

As Kuk (2006) noted at the close of his study, upon which much of this current research
is based:
The structural measures of this study offer a practical way of deriving proxy measures of process effectiveness in knowledge sharing. For example, managers can benchmark the proxy measures with other performance indicators such as product quality and customer satisfaction, and use the benchmarking findings to compute their internal “optimum concentration density” curve, by plotting the performance indicators against the structural measures. Through this, managers can identify the optimum point of structural concentration and use it to determine an acceptable range before making any attempts to intervene (p. 1040).

Divining individual (to say nothing of group-level) participation in groups has always been an unreliable, time-consuming, and arduous task, and nearly impossible to do with any level of efficiency, confidence and accuracy. Visible Effort eliminates the guesswork instantaneously and provides participation levels that are of a mathematical exactness. None of this is magic, but simply algorithmic processes that are theoretically rooted in information theory and enabled through the latest in emerging collaborative technology – a wiki.

Most fundamentally, the novel applications of theory and the technology to carry it out addressed by this paper could have far-ranging and dramatic impact on education, business and many other settings, processes and applications. Emerging social technologies like wikis are presently transforming human communication and knowledge production as we know it, and these tools and their respective utilities such as Visible Effort have the power to change the world through new forms of collaboration and learning.