

Työn vaatimukset prosesseina



Tiina Kalliomäki-Levanto
Pekka Varje
Marja Käsälä
Ilkka Kivimäki
Olli Haavisto

Työn vaatimukset prosesseina

Koneoppimismenetelmät ja teoriat tunnistamisen apuna

Tiina Kalliomäki-Levanto

Pekka Varje

Marja Känsälä

Ilkka Kivimäki

Olli Haavisto

Työterveyslaitos

Työkyky ja työurat

PL 40

00032 Työterveyslaitos

www.ttl.fi

© 2023 Työterveyslaitos ja kirjoittajat

Hanke on toteutettu Työsuojelurahaston tuella.

Tämän teoksen osittainenkin kopiointi on tekijänoikeuslain (404/61, siihen myöhemmin tehtyine muutoksineen) mukaisesti kielletty ilman asianmukaista lupaa.

ISBN 978-952-391-077-5 (PDF)

Tiivistelmä

Raportissa kerrotaan Työsuojelurahaston rahoittamasta hankkeesta ”Organisaatiodatan hyödyntämisen luotettavuuden lisääminen teorian ja tekoälyn avulla”. Hanke on teoreettismetodologinen, ja sen tavoitteena oli kehittää prosessimallia, jolla organisaatiot voivat seurata työhyvinvoinnin kehittymistä. Aineistona käytettiin aiemmin TYÖ2030 -hankkeen puitteissa ammattikorkeakouluista kerättyä opettajien anonymisoitua Moodle-dataa. Moodle on virtuaalinen oppimisympäristö, jota opettajat käyttävät opetuksen suunnittelua, toteutusta ja arviointia varten. Moodlen käytöstä jää jälki (trace data), joka sisältää tiedon toimijasta (anonymisoitu), ajankohdasta (päivätasoinen tieto) sekä kohteesta (target) ja toiminnasta (action), esim. message_sent.

Hankkeessa tuotetussa ensimmäisessä artikkelissa osoitimme, miten järjestelmien käytössä syntyvä jälki, trace data, voidaan kiinnittää prosessiteorian viitekehykseen (Langley ym., 2013). Näin ollen tulee mahdolliseksi havaita erilaista toiminnan ajallista etenemistä eli prosesseja ja tehdä päätelmiä työhön liittyvien prosessien mahdollisesti kuormittavista ajankohdista. Päätelmien apuna käytimme Karasekin työn vaatimukset-hallinta -mallia (1979). Toisessa artikkelissa jatkoimme analyysiä tarkemmalle tasolle ja käytimme laskennallisena menetelmänä klusterointia, jolla tunnistimme kuusi erilaista prosessia erilaisin vaatimuksin tehdä opetustyötä keväällä 2019 ja 2020.

Elovainio ym. (2022) ehdottavat, että psykososiaalisten työolosuhteiden parantaminen tulisi kohdistaa työn kuormittaviin vaatimuksiin, jolla on ajallisesti myöhemmin myönteinen vaikutus muihin psykososiaalisiin työolosuhteisiin. Tämä edustaa uutta näkökulmaa, kun tähän mennessä huomio on kiinnitetty työn hallinnan kehittämiseen. Uutena psykososiaalisena tekijänä tunnistimme proaktiivisuuden (ennakointi, varautuminen tulevaan), jonka Moodle mahdollistaa.

Lähetimme kummatkin käsikirjoitukset arvioitavaksi tieteellisiin lehtiin ja luottamuksellisena tietona Työsuojelurahastolle osana loppuraportointia. Päätulosten visualisointeja eli toiminnan ajallista etenemistä ei raportoida tässä raportissa, koska visualisoinnit ovat artikkelikäsikirjoitusten keskeisinä tuloksina ja tulevat julkiseksi vasta kun artikkelit julkaistaan. Luottamuksellisena koulutusmateriaalina muokattuja visualisointeja voidaan käyttää.

Abstract

The report presents the results from the research project "Using theory and artificial intelligence to Improve the reliability of the utilization of organizational data". The research project utilized a theoretical-methodological approach and aimed at developing a process model that organizations can use to monitor the developments in occupational wellbeing. The research materials consisted of anonymized teacher-generated Moodle-data collected from two Finnish universities of applied sciences, previously gathered for the WORK2030-project. Moodle is virtual educational platform used by teachers for planning and conducting courses as well as for evaluating the students. All operations conducted with Moodle leave a trace that includes information on the actor (anonymized for research), timing (at daily level), target and action, i.e. message_sent.

In the first research article related to the project we show how the trace data generated by information systems can be attached to the framework of process theory (Langley et al. 2013). This enables the tracking of the temporal progressions of activities, namely processes, and allows for the identification of the time points at which the processes constitute of potential source of job strain. In our conclusions we utilize the demand-control model developed by Karasek (1979). In the second article, we continued the analysis to a more detailed level and used clustering as a computational method, which we used to identify six different processes with different job demands to do teaching work in the spring of 2019 and 2020.

As suggested by Elovainio et al. (2022), improving psychosocial work environment should focus on those strenuous job demands that have the potential to be efficient targets of workplace interventions. This argument represents a novel approach since earlier contributions have focused on developing job control. As a new psychosocial factor, we identified proactivity (anticipation, preparation for the future), which Moodle enables.

The manuscripts of both our articles were submitted to two scientific journals for review. Both manuscripts have also been confidentially submitted to the Finnish Work Environment Fund as a part of the final report.

Visualizations of the main results related to the temporal progression of activities are not published in this report because they represent the main outcomes of the unpublished manuscripts. The visualizations will become available with the publication of the manuscripts and meanwhile they can be utilized as confidential educational material.

Sisällys

Tiivistelmä.....	3
Abstract	4
Sisällys.....	5
1 Johdanto.....	6
2 Aineisto ja aiheet (topics).....	8
2.1 Aineisto	8
2.2 Aiheet.....	8
3 Artikkel 1: Digitaalisen trace datan kiinnittäminen prosessiteoriaan aihemallinnuksen avulla.....	11
3.1 Prosessiteoria.....	11
3.2 Tulokset ja pohdinta	12
4 Artikkel 2: Aihemallinnuksella tunnistetut työn vaatimukset dynaamisissa klustereissa.....	15
4.1 Työn vaatimukset.....	15
4.2 Tulokset ja pohdinta	16
5 Päätelmät.....	20
Lähteet.....	21

1 Johdanto

”Organisaatiodatan hyödyntämisen luotettavuuden lisääminen teorian ja tekoälyn avulla” -hankkeen lähtökohtana on digitaalisuus. Kun digitaalisia järjestelmiä käytetään yhä enemmän organisaatioissa, niin yhä enemmän digitaalista trace dataa tallentuu järjestelmien lokitietoihin. Kun järjestelmiä käytetään tehtävien suorittamiseen, samalla muodostuu informaatiota työn ja organisaation käytännöistä. Järjestelmiin kertyvän tiedon hyödyntämiseen työelämän kehittämisessä on kohdistettu kolmenlaista epäilyä. Ensiksi tutkimuksen piirissä on epäilyjä sille millaista ilman tutkijan teoriaa ja menetelmää kertyvä tieto on, ja toiseksi mikä on tiedon suhde olemassa oleviin teorioihin. Kolmas epäily on se, että hyödynnetäänkö tietoa henkilöstön työhyvinvoinnin hyväksi. Näitä epäilyjä hälvennetään kahden tieteellisen artikkelin avulla.

Aineistona käytettiin aiemmin TYÖ2030 -hankkeen puitteissa ammattikorkeakouluista kerättyä ja anonymisoitua opettajien Moodle-dataa. Moodle on virtuaalinen oppimisympäristö, jota opettajat käyttävät opetuksen suunnittelua, toteutusta ja arviointia varten. Moodlen käytöstä jää jälki (trace data), joka sisältää tiedon toimijasta (anonymisoitu), ajankohdasta (päivätasoinen tieto) sekä kohteesta (target) ja toiminnasta (action), esim. message_sent. Ensimmäisessä artikkelissa osoitetaan, miten järjestelmien käytössä syntyvä trace data voidaan kiinnittää prosessiteorian viitekehykseen (Langley ym., 2013). Näin ollen tulee mahdolliseksi havaita erilaista toiminnan ajallista etenemistä eli prosesseja ja tehdä päätelmiä työhön liittyvien prosessien mahdollisista ajankohdista ja ajanjaksoista. Päätelmien apuna käytetään Karasekin työn vaatimukset-hallinta -mallia (1979).

Toisessa artikkelissa jatkettiin analyysiä tarkemmalle tasolle ja käytettiin laskennallisena menetelmänä klusterointia, jolla tunnistetaan kuusi prosessia erilaisin vaatimuksin tehdä opetustyötä keväällä 2019 ja 2020. Elovainio ym. (2022) ehdottavat, että psykososiaalisten työolosuhteiden parantaminen tulisi kohdistaa työn vaatimuksiin, jolla on ajallisesti myöhemmin myönteinen vaikutus muihin psykososiaalisiin työolosuhteisiin. Tämä edustaa uutta näkökulmaa, kun tähän mennessä huomio on kiinnitetty työn hallinnan kehittämiseen. Uutena psykososiaalisena tekijänä tunnistettiin proaktiivisuus, jonka Moodle mahdollistaa. Proaktiivisuudella tarkoitetaan ennakoitua ja tulevaan varautumista, mikä omalta osaltaan voisi edistää työn vaatimusten hallintaa.

Artikkeleissa menetelminä ovat datalähtöiset koneoppimismenetelmät, jotka tunnistavat tietoa datasta ja tunnistettu tieto voidaan liittää teoreettisiin viitekehyksiin. Näin ollen organisaatiodatan hyödyntämisen luotettavuus lisääntyy. Kun on

mahdollisuus tunnistaa uudella tavalla työn vaatimuksia ja mahdollisesti vaikuttaa niihin, voidaan edistää henkilöstön työhyvinvointia.

Kyselylomakkeet ovat olleet käytetyin menetelmä hankkia tietoa työstä kuten työn vaatimuksista ja hallinnasta. Kyselyitä on toistettu esimerkiksi vuoden tai kahden välein ja saatu pistemäisesti tietoa pidemmältä ajalta. Näin saatu tieto ei kuitenkaan ole jatkuvaa kuten trace datasta tunnistettava prosessimainen tieto on. Kyselyiden avulla saadaan koettua tietoa työstä, kun trace data tuottaa realistista ja käytännöllistä tietoa työstä.

2 Aineisto ja aiheet (topics)

2.1 Aineisto

Tutkimusaineisto koostuu opettajien luomasta Moodle-trace datasta kahdesta ammattikorkeakoulusta. Data sisältää 11 (org. A) ja 5,7 (org. B) miljoonaa riviä trace dataa 90 viikon ajalta vuosilta 2019–2020 (n. 400 henkilöä org. A ja n. 300 henkilöä org B). Yksittäinen trace koostuu neljästä osasta: henkilön tunnistetiedosta, päivämäärästä ja kohteesta (target) sekä toiminnasta (action), jotka muodostavat yhdessä target_action parin kuten user_graded. Taulukossa 1 on esimerkkejä trace datasta.

Taulukko1: Satunnaisotanta org. A:n Moodle trace datasta

teacher id	date	target > action
<id_40>	17.6.2019	course_module > viewed
<id_265>	17.12.2019	course > viewed
<id_352>	3.6.2020	course_module > updated
<id_282>	26.8.2020	course > viewed
<id_395>	18.2.2020	grading_form > viewed
<id_228>	27.1.2020	course > viewed
<id_127>	26.8.2019	user_list > viewed
<id_283>	10.5.2019	user > graded

Organisaatiot A ja B, jotka omistavat Moodle-datan, anonymisoivat tiedot GDPR:n mukaisesti. Anonymisointi tarkoitti sitä, että henkilöille annettiin satunnainen tunniste, jonka yhteys henkilön tunnistetietoon hävitettiin. Lisäksi aikaleima tuli sisältää vain päiväkohtaisia tietoja.

2.2 Aiheet

Kummankin organisaation aineistoista tunnistettiin aihehallinnuksella (Blei ym. 2003) 17 aihetta (topics). Aihemallinnus on ohjaamaton koneoppimismenetelmä, joka käyttää aineistona dokumentteja, joissa on sanoja. Hankkeen yksittäisen dokumentin muodostivat opettajan yhden viikon target_action-parit, jotka muodostivat joukon target_action sanoja. Jos opettajia on n. 400 ja viikkoja 90, saadaan n. 36 000 dokumenttia.

Aihemallinnuksessa algoritmi vertailee dokumentteja keskenään dokumentissa olevan sanojen muodostaman jakauman perusteella. Sanajakaumassa voi 1–2 sanaa

dominoida eli sanat esiintyvät usein dokumentissa ja muutamat sanat voivat esiintyä toistuvasti monissa dokumenteissa yhdessä. Dokumentit, joiden sanajakaumat ovat samanlaisia rytmittyvät yhdeksi aiheeksi. LDA tunnisti erilaisia aiheita 17 kummassakin organisaatiossa. Aiheet muodostavat jakauman ja niiden todennäköisyydet summautuvat ykköseen koko aineistossa ja yksittäisessä viikkodokumentissa. Yksittäisen aiheen nimeämiseen käytetään sanajakaumasta niitä sanoja, jotka dominoivat ja joilla on siis suurin todennäköisyys.

Lisäanalyysiä varten valittiin 11 aihetta. Aiheet, joissa aktiivisuus oli lähes nollassa suurimman osan viikkoista ja joidenkin yksittäisten viikkojen huiput olivat korkeat, jätettiin pois. Poisjätetyt nimettiin teemaksi muut, jota käytettiin toisen artikkelin visualisoinneissa. 11 aihetta ryhmiteltiin tutkijan toimesta neljään taulukossa 2 lueteltuun teemaan, ja aihe #16 sisällytettiin sellaisenaan, koska tulkitsemme sen edustavan opetustyön keskeistä osaa: loppuarviointia. Sillä voi myös olla rooli aikapaineen aiheuttajana.

Taulukko 2: Tutkijan toimesta muodostetut teemat (org. A).

Teemat	Teemojen aiheet ja aiheessa dominoivat sanat (target_action parit)
Arviointi	Topic #9: grading_form_viewed submission_graded Topic #17: grading_table_viewed submission_status_viewed
Kurssit	Topic #4: course_module_updated Topic #10: course_bin_item_created course_module_deleted Topic #12: course_section_created course_section_updated
Edistymisen seuranta	Topic #3: attempt_reviewed question_manually_graded report_viewed Topic #15: attempt_viewed edit_page_viewed
Tilanteen seuranta	Topic #1: dashboard_viewed Topic #7: discussion_viewed Topic #13: user_list_viewed
Lopullinen arviointi	Topic #16: submission_graded user_graded

Yksittäinen viikkodokumentti teemajakaumatiedoin muodostaa tulosaineiston, jota käytetään kahdessa artikkelissa eri tavoin. Ensimmäisessä artikkelissa (luku 3) visualisoidaan jokaisen teeman opettajien keskimääräinen todennäköisyyksien (intensiteetin) kehittyminen viikosta toiseen samaan kuvaan, josta voidaan havaita toiminnan intensiteetin ajallinen eteneminen sekä intensiteettien keskinäisten suhteiden muutokset. Tällä tavalla on mahdollista tunnistaa Langelyn ym. (2013) tarkoittama prosessi työn vaatimuksista (Karasek 1979).

Toisessa artikkelissa (luku 4) jokaiselle opettajalle jokaiselle kevään viikolle saadaan teemajakauma ja näiden jakaumien avulla voidaan tehdä klusterointia. Kahden ammattikorkeakoulun opettajien viikoittaiset teemajakaumat yhdistettiin yhdeksi

aineistoksi, jolloin voidaan tulkita yleisellä tasolla opetustyötä ammattikorkeakouluissa. Klusterointia varten ensin määriteltiin opettajan jakauman ero kevään viikoilla kaikkiin muihin opettajiin. Näin saadun etäisyysmitan avulla klusteroitiin opettajat ohjaamattomalla koneoppimismenetelmällä (Landau ym. 2011) ja saatiin kuusi erilaista klusteria kevään työskentelystä, joita voidaan tulkita työn vaatimukset-hallinta mallin avulla (Karasek 1979).

3 Artikkel 1: Digitaalisen trace datan kiinnittäminen prosessiteoriaan aihehallinnuksen avulla

3.1 Prosessiteoria

Aikaleimatun tiedon (luku 2) avulla on mahdollista määritellä ajallisia elementtejä, kuten tapahtumia, sekvenssejä (tapahtumat seuraavat toisiaan ajallisesti peräkkäin), vaiheita ja jaksoja. Teoriassa digitaalisen trace datan avulla on mahdollista tarkastella toimintaa prosessina joustavasti minuuttien, tuntien, päivien, viikkojen, kuukausien tai vuosien tasolla. Käytännössä viikkotason tieto on riittävä työprosessien tunnistamiselle tietointensiivisessä työssä.

Toisaalta teoria organisaatioiden prosesseista alkoi kehittyä jo kauan ennen digitaaliseen trace dataan perustuvien lähestymistapojen ilmaantumista. Ann Langley ja hänen kollegansa ovat kehittäneet teoriaa organisaation prosesseista vuodesta 1999 lähtien. He määrittelevät prosessin sisältävän vastaavia ajallisia elementtejä kuin trace datan kanssa työskentelevät: tapahtumia, sekvenssejä, vaiheita ja jakso. On huomattava, että prosessin lähtökohta, tapahtuma, eroaa trace dataan ja prosessiteoriaan perustuvissa lähestymistavoissa toisistaan. Langleyn prosessikäsitteen mukaan tapahtuma voi sisältää esimerkiksi huonon vuoden, fuusion, päätöksen, kokouksen, keskustelun tai kädenpuristuksen. Trace data -lähtöisissä prosesseissa tapahtuma on selkeästi määritelty (ajankohta, toimija, kohde, teko). Prosessiteorian kehittäjät hankkivat tutkijavoimin tietoa sekvensseistä, vaiheista ja jaksoista eli haastattelemalla, havainnoimalla tai dokumenttien avulla. Näin ollen ajankohdat tulee etukäteen valita ja kaikki tiedon hankinta on rajallista inhimillisin voimin tehtävää, kun trace data mahdollistaa erilaisten aikaikkunoiden ja sekvenssien kokeilut.

Prosessiteorian ajallisten elementtien ja trace data -lähtöisen lähestymistavan välillä on yhtäläisyyksiä, ja lisäksi molemmat tunnistavat prosesseja datalähtöisesti (Langley et al., 2013; Pentland et al., 2021). Jos nämä yhtäläisyydet voidaan tunnistaa yksityiskohtaisemmin, epäilykset trace datan luotettavasta käytöstä voivat hälventyä.

Prosessitutkimus tarkastelee empiirisesti kehittyviä ilmiötä eli toiminnan ajallista etenemistä, minkä avulla voidaan selittää ja ymmärtää ilmiötä. Ilmiöitä ei siis selitetä kahden muuttujan välisellä korrelaatiolla vaan tarkemmin prosessin ilmaantumisenä (becoming) ja prosessien välisinä yhteyksinä. Prosessitutkimus perustuu kahdelle erilaiselle sosiaalisen maailman ontologialle. Ensimmäinen väittää, että entiteetit koostuvat asioista, joissa prosessit edustavat muutosta asioissa, jotka tiedetään etukäteen (perustuu substantiiviseen metafysiikkaan, Langley et al., 2013).

Toinen ontologia viittaa entiteetteihin, jotka muodostuvat prosesseista, joissa asiat ovat jatkuvasti muuttumistilassa (becoming) (prosessimetafysiikka, Langley et al., 2013). Dynaamiset prosessit ovat vakauden ja muutoksen taustalla. Prosessinäkökulmasta katsottuna tulokset, kuten tietynä ajankohtana mitattu organisaation suorituskyky, ymmärrettäisiin kuitenkin yleensä ohimenevinä havaintona jatkuvassa toiminnan virrassa (Langley et al., 2013). Trace datan avulla on mahdollista löytää myös muuttumistilassa olevia prosesseja eikä vain etukäteen tiedettyjä yksittäisiä tapahtumia ja niiden prosesseja.

Prosessitutkimuksella pyritään ymmärtämään ensinnäkin, miten asiat kehittyvät ajan myötä, ja siksi prosessidata koostuu tiedosta siitä, mitä tapahtui ja kuka teki mitä ja milloin, ja toiseksi, miten tapahtumat järjestyvät ajan myötä ja muodostavat sekvenssejä. (Langley, 1999). Nämä kaksi ymmärrystä ovat myös osa prosesseja, jotka löydetään automaattisesti eri tavoin trace datasta (Pentland ym. 2021).

Prosessiteoria on kehitetty tutkijoiden itse toteuttaman tiedonkeruun avulla. Tietoa on kerätty pistemäisesti rajatun ajanjakson aikana hyödyntäen grounded teoriaa, mikä vaatii paljon inhimillistä resurssia. Tuloksena on saatu käsitteitä, jotka viittaavat toimintoihin, joilla on vaihtelua aikapisteiden mukaan (Langley et al., 2013). Tämä kaltainen analyysi on mahdollista myös aihemallinnuksen avulla (Lindberg, 2020).

Visualisoinnin avulla on mahdollista havaita muutokset prosessien välisissä suhteissa ajan myötä. Lisäksi kyky tulkita useiden prosessien visualisointeja olemassa olevan teorian avulla mahdollistaa ilmiöiden uuden ymmärtämisen (Langley et al., 2013). Tällainen päättely on mahdollista myös trace data -lähtöisillä prosesseilla.

3.2 Tulokset ja pohdinta

Visualisointien avulla voidaan havaita, että työ etenee prosessina melko samalla tavalla kahdessa ammattikorkeakoulussa. Emme julkaise visualisointeja tässä raportissa, koska ne ovat keskeinen tulos käsikirjoituksissa, jotka lähetettiin tieteellisiin lehtiin. Kuvailimme sanallisesti visualisointeja.

Moodle-toiminnan visualisointeja tarkasteltaessa tulee huomioida, että toiminnan suhteellinen osuus Moodlessa ei suoraan kerro aiheeseen sisältyvän tekemisen vaatimuksia. "Submission_graded user_graded" -toiminto saattaa vaatia paljon resursseja henkilöltä, muilta ja koko organisaatiolta Moodlen ulkopuolella. Toisaalta "course_module_updated" -toiminto voi perustua jo olemassa olevaan kurssiin ja päivitys on rutiinitoimintaa eikä vaadi niin paljon resursseja järjestelmän ulkopuolelta. Jotkut target-action-parit voivat tarkoittaa yhtä klikkausta ja toiset useita tunteja työtä.

Siksi jonkin aiheen aktiivisuuden lievä lisääntyminen, jonka suhteellinen osuus on pieni muihin toimintoihin verrattuna, voi kuitenkin olla merkki vaatimuksen kasvusta. Opetussektorilla voi olla yksi ilmeinen vaatimus, nimittäin aikapaine muutamaa viikkoa ennen kesää.

Arviointiin, kursseihin, tilanteen ja edistymisen seurantaan ja lopulliseen arviointiin liittyvien prosessien väliset suhteet muuttuivat kevään 2019 ja 2020 aikana. Arvioinnin intensiteetti kasvaa molempina vuosina kesää kohti, jolloin huippu on viikoilla 21–22. Työskentelyn intensiteetin lisääntyminen voi vastata Karasekin mallin (1979) työn vaatimuksia, joita on tutkittu esim. työhyvinvointikyselyillä esimerkiksi osioilla "onko työtä paljon" ja "riittääkö aika". Kurssien intensiteetti laskee kesää kohti vuosina 2019 ja 2020. Tilanneseurannan intensiteetti pysyy molempina vuosina melko samana viikkojen ajan ja hieman laskee kesää kohti. Lopullisen arvioinnin intensiteetti pysyy melko samana viikkojen ajan molempina vuosina, hieman laskien kesää 2020 kohti.

Vuonna 2019 "edistymisen seurannan" intensiteetti pysyy melko samana viikkojen ajan, mutta vuonna 2020 intensiteetti kasvaa selvästi viikon 10 jälkeen kesää kohti saavuttaen huippunsa viikolla 20. Viikko 10 vuonna 2020 oli COVID-19-viikko, kun työskentely kaikilla sektoreilla, myöskin opetuksessa, siirtyi mahdollisimman paljon virtuaaliseksi ja digitaaliseksi.

Kun opetus muuttui etänä toteutettavaksi ja kasvokkain tapahtuva vuorovaikutus oppilaiden edistymisen tietolähteenä katosi, tarvittiin uusia keinoja edistymisen seuraamiseen. Opiskelijoiden edistymisen seurantatietoa löytyi Moodlesta. Tämä opiskelijoiden edistymisen seurantaan viittaava muutos opettajien aktiivisuuden intensiivisyydessä voi olla esimerkki prosessiteorian mukaisesta "becoming" prosessista (ks. Langley ym., 2013) – edistymisen seuranta voi olla merkki uudesta kehittyvästä tavasta, jolla opettajat voivat tukea opiskelijoiden edistymistä Moodlesta.

Yllä on kuvailtu prosesseja yleisellä tasolla. Tarkastelimme myös prosessien välisiä yhteyksiä. Tulosuunnitukseksi valittiin "lopullinen arviointi" (Topic #16: submission_graded user_graded) painottaen erityisesti kevätlukukauden viimeistä viikkoa (viikko 22). Eletään kevätlukukauden kiireisintä aikaa, jolloin myös "lopullinen arviointi" saa korkeimmat intensiteettiänsä, saaden kuitenkin tutkimusjaksolla matalia arvoja verrattuna muihin prosesseihin. Testattavana oletuksena oli, että jos viikon 22 loppuarvioinnin intensiteetti on opettajaryhmällä alhainen, ehkä he voivat jakaa työmääränsä tasaisemmin kevään aikana. Tämän pitäisi johtaa teeman intensiteetin vähäisempään vaihteluun viikosta toiseen (vähemmän huippuviikkoja), kun taas opettajilla, joilla on paljon loppuarviointityötä viikolla 22, intensiteettien vaihtelua (huippuviikkoja) voisi olla enemmän.

Oletuksen selvittämiseksi opettajat jaettiin kahteen ryhmään kiireisimmän viikon (22) loppuarvioinnin intensiteetin mukaan. 'Ryhmä matala' koostui opettajista, joilla ei ollut juurikaan loppuarviointia viikon 22 aikana, ja 'ryhmän korkea' muodostivat loput opettajat. Ryhmät perustettiin kummastakin vuodesta erikseen ja saatiin 4 kuvaa (v. 2019 matalan ja korkean ryhmän opetustyön prosessit ja v. 2020 vastaavasti)

Tarkasteltaessa kaikkia neljää kuvaa samanaikaisesti näyttää siltä, että molemmat 'matala' ryhmät ja toisaalta molemmat 'korkea' ryhmät ovat samanlaisia. 'Korkea' ryhmissä on kurssien huippuviikkoja vuoden alussa ja arvioinnin ja loppuarvioinnin intensiteetin tasaisesti nouseva jakso noin viikosta 16 viikkoon 22. Nämä ominaisuudet puuttuvat 'matala' ryhmistä. Voimme ehdottaa, että 'matala ja 'korkea' ryhmät ovat erilaisia: Työskentely 'matala' ryhmässä on vakaampaa ja voi siksi olla vähemmän kuormittavaa kuin 'korkea' ryhmässä. Nyt on mahdollista tunnistaa työn sisällöstä työn kuormittavia vaatimuksia sen sijaan, että tietoa vaatimuksista hankitaan kyselyillä koettuna tietona kiireestä, työn määrästä ja ajan riittämisestä kuten Karasekin (1979) mallia pääsääntöisesti on tähän mennessä sovellettu.

4 Artikkelit 2: Aihemallinnuksella tunnistetut työn vaatimukset dynaamisissa klustereissa

4.1 Työn vaatimukset

Karasekin (1979) työn vaatimukset-hallinta -malli ennustaa, että henkinen paine/rasitus (strain) johtuu työn vaatimusten ja työn hallinnan (kontrollin) vuorovaikutuksesta. Mallin mukaan matalan hallinnan, kuten vähäinen mahdollisuus osallistua päätöksentekoon, ja kovien työvaatimusten yhdistelmä liittyy henkiseen paineeseen, joka pitkään jatkuessaan muuttuu haitalliseksi stressiksi. Henkisen paineen vähentäminen voidaan saavuttaa suunnittelemalla työprosesseja uudelleen siten, että hallintaa voidaan lisätä vähentämättä työn vaatimuksia, kuten työmäärää, jotka puolestaan liittyvät organisaation tuottotavoitteisiin. Karasekin ehdotuksen mukaisesti mallia vuodesta 1979 lähtien testanneet lukuisat interventiotutkimukset ovat keskittyneet pääasiassa mallin hallinta-ulottuvuuteen eivätkä vaatimuksiin.

Kuitenkin tuoreessa julkaisussa Elovainio ym. (2022) ehdottavat interventioiden painopisteen siirtämistä vaatimus-ulottuvuuteen. Psykososiaalista työympäristöä koskevassa analyysissään he totesivat työn vaatimukset yhdeksi keskeisistä tekijöistä, joka voisi olla arvokas kohde työpaikan interventioihin. Työn vaatimusten, parantaminen johtaisi positiivisen vaikutuksen leviämiseen ajan myötä myös psykososiaalisen verkoston muihin osiin, kuten tiimin ilmapiiriin tai kokemukseen ponnistelun ja palkkion tasapainosta. Lisänäyttö työn vaatimusten dynamiikasta voisi helpottaa tietoon perustuvia päätöksiä siitä, kenelle, milloin ja miten toimia työympäristön ja työhyvinvoinnin parantamiseksi.

Työn vaatimusten asteikon rakentamisen tavoitteena on ollut mitata työkuormituksen suorittamiseen liittyviä psykologisia stressitekijöitä. Aikaisemmissa Karasekin (1979) mallia hyödyntävissä tutkimuksissa työn vaatimuksia on analysoitu enimmäkseen tukeutumalla kyselyihin, jotka perustuvat työntekijöiden omiin subjektiivisiin käsityksiin työstään. Nämä kyselyt ovat perustuneet seuraaviin Karasekin määrittelemiin työn vaatimusten ominaisuuksiin: työskentelee nopeasti, työskentelee kovasti, on paljon työtä, aika ei riitä, liiallinen työmäärä, ei aikaa saada valmiiksi ja ristiriitaiset vaatimukset.

Psykososiaalisista tekijöistä, kuten vaatimuksista ja hallinnasta, kyselylomakkeilla kerätyn koetun tiedon pätevyyttä on arvioitu altistumistutkimuksessa, jossa työn vaatimuksia koskevan koetun tiedon luotettavuus oli heikko. Tutkijat ehdottavat, että työn vaatimuksista tarvittaisiin tarkempaa tietoa, mutta useimmat tutkimukset hyödyntävät silti koettua tietoa, koska tarkempaa tietoa ei ole saatavilla.

Tarve saada objektiivisempaa tietoa työn vaatimuksista, kuten tarkempia analyysejä työstä ja tallenteita organisaation toiminnasta, on nostettu esiin. Harvinaisessa esimerkissä Rau ym. (2010) käyttivät objektiivista tietoa työn vaatimuksista tutkiessaan työn vaatimusten ja masennuksen välistä yhteyttä. Tieto työn vaatimuksista perustui Task Diagnosis Survey -menetelmään (Rau ym., 2010). Tuota menetelmää voidaan pitää "objektiivisena", koska työanalyysin asiantuntijat arvioivat työn ominaisuuksia riippumatta siitä, miten työntekijä kokee työtehtävänsä. He pystyivät osoittamaan, että objektiivisesti määritellyt vaatimukset liittyivät merkittävästi masennukseen.

Spector ja Pindek (2016) ovat huomauttaneet, että digitaalinen big data, trace data, voisi olla uusi tietolähde työstä ja työympäristöstä organisaatioissa. Datalähtöisillä lähestymistavoilla, joissa käytetään suuria datalähteitä, voi olla potentiaalia tarjota arvokkaita oivalluksia moniin organisaation ilmiöihin. Näitä lähestymistapoja voitaisiin käyttää hahmojen tunnistamiseen big datasta ja tärkeiden seurausten, kuten sairauden tai vamman, ennustamiseen.

Sen lisäksi, että tarvitaan objektiivisempaa tietoa työn vaatimuksista, tarvitaan myös dynaamista ja prosessimaista tietoa, joka voisi paremmin saada esiin nykyajan työn vaatimukset. Prosessien tutkimusta varten tarvitaan ajallisia sekvenssejä, joiden tulee sisältää alku, loppu ja välivaiheet. Vaikka kyselyillä hankittu tutkimustieto on usein pitkittäistä, se on myös staattista, kerättyinä tiettyinä yksittäisinä ajankohtina. Kyselydataa käyttämällä ei ole mahdollista löytää työprosessia aikaleimatun trace datan mahdollistamalla tavalla. Trace datalla on aikaleima, jonka avulla voidaan havaita toiminnan ajallinen eteneminen prosesseina (vrt. Langley ym. 2013). Trace datan avulla on mahdollista saada objektiivisempaa tietoa sekä työtoiminnan käytännön sisällöstä että niiden ajallisesta etenemisestä.

Edellä luvussa 2 on esitetty kahden ammattikorkeakoulun opettajien luoma Moodle-trace data, josta aihemallinnuksen avulla tunnistettiin aiheita, joista muodostettiin kuusi teemaa kertomaan opetustyöstä: arviointi, kurssit, tilanteen ja edistymisen seuranta, lopullinen arviointi ja muu toiminta. Jokaiselle opettajalle saatiin jokaiselle viikolle teemajakauma omasta työskentelystä ja aineistot voitiin yhdistää yhdeksi kokonaisuudeksi. Opettajat klusteroitiin datalähtöisesti ja saatiin kuusi klusteria.

4.2 Tulokset ja pohdinta

Kuuden klusterin visualisointien avulla voidaan havaita, että työ etenee prosessina osittain eri tavoin klustereissa. Emme julkaise visualisointeja tässä raportissa, koska ne

ovat keskeinen tulos käsikirjoituksessa, joka lähetettiin tieteelliseen lehteen arvioitavaksi. Tulkitsemme sanallisesti visualisointeja taulukon 3 avulla.

Taulukko 3. Kuusi klusteria keväällä 2019 ja 2020 nimettiin opetustyön teemojen keskinäisen suhteen muutoksen avulla. Samansuuntaista tietoa sisältävät klusterit vuosina 2019 ja 2020 ovat samalla rivillä. Vuoden 2020 kuudes klusteri eroaa eniten kaikista muista ja on siksi omalla rivillään.

Klusteri	Kevät 2019 (viikot 1-22)	Klusteri	Kevät 2020 (viikot 1-22)
1	Arvioinnin (arviointi+lopullinen arviointi) osuus kasvaa kesää kohden, kurssisiin liittyvä tekeminen vähenee ja edistymisen seuranta on tasaisesti vähän: "Arviointi lisääntyy"	1	Arvioinnin (arviointi+lopullinen arviointi) osuus kasvaa kesää kohden, kurssisiin liittyvä tekeminen vähenee ja edistymisen seuranta kasvaa hieman kesää kohden. "Arviointi lisääntyy ja edistymisen seuranta lisääntyy vähän"
2	Arvioinnin osuus kasvaa kesää kohden ja tilanteen seuranta vähenee: "Arviointi lisääntyy"	3	Arvioinnin osuus kattaa noin neljänneksen toiminnasta ja lisääntyy kesää kohden, kun muiden teemojen osuus vähenee. "Arviointi lisääntyy"
3	Arvioinnin osuus pysyy samalla tasolla viikosta 4 lähtien kesää kohden ja arvioinnin osuus kaikkiaan on suurin tässä klusterissa. "Enimmäkseen arviointia"	2	Arvioinnin osuus pysyy miltei samalla tasolla viikolta 4 lähtien päätyn vähäiseen laskuun ennen kesää ja arvioinnin osuus on suurin kuudesta klusterista. Edistymisen seuranta on tasaisen vähäisellä osuudella. "Enimmäkseen arviointia vähäisellä edistymisen seurannalla"
5	Arvioinnin osuus vähenee viikkoa 10 kohden ja sen jälkeen kasvaa kesää kohden ollen viikolla 22 osuudeltaan suurin kuuden klusterin joukossa. "Arviointi lisääntyy korkeimpaan osuuteen viikolla 22"		
4	Arvioinnin osuus pysyy samalla tasolla viikosta 10 eteenpäin kohti kesää ollen noin puolet koko aktiivisuudesta, kun toinen puoli sisälsi muiden teemojen aktiivisuutta. "Puolet arviointia ja puolet muuta tekemisen teemoja"	4	Arvioinnin osuus kattaa noin kolmanneksen aktiivisuudesta ja lisääntyy kesää kohden, kun edistymisen seuranta puolestaan lisääntyy viikosta 11 lähtien voimakkaasti kesää kohden ja aktiivisuusosuus muissa teemoissa vähenee. "Edistymisen seuranta lisääntyy voimakkaasti kesää kohden"

6	Arvioinnin osuus vaihtelee viikosta toiseen päätyen vähentymään päin ennen kesää. Vastaava vaihtelua ilmenee edistymisen seurannassa, jolla on selkeä rooli vain tässä klusterissa. "Arviointia ja edistymisen seurantaa"	5	Arviointi pysyy miltei samalla tasolla viikosta 10 eteenpäin, kun edistymisen seuranta puolestaan lisääntyy ja aktiivisuus muissa teemoissa vähenee kesää kohden. "Edistymisen seuranta lisääntyy kesää kohden"
		6	Arviointi kattaa pienen osuuden kevään aktiivisuudesta ja edistymisen seuranta samoin, jotka pysyvät samalla tasolla kevään viikoilla. Suurempi osuus aktiivisuudesta on muilla teemoilla ja osuudet vaihtelevat viikosta toiseen. "Muuta aktiivisuutta kuin arviointia tai edistymisen seurantaa"

Poimiaksemme yksityiskohtaista tietoa interventiota varten jatkamme klustereiden tulkintaa sisällön ja prosessien muutoksen näkökulmasta. "Enimmäkseen arviointia"-klusterien (2019:3, 2020:2) sisältö voi vastata "paljon työtä" osiota Karasekin mallissa (1979) ja siten edustaa kuormittavia työn vaatimuksia. Kesää kohti nouseva aktiivisuus näkyy vuonna 2019 klustereissa 1 ja 2 ja vuonna 2020 klustereissa 1 ja 3, joissa arviointiin liittyvä tekeminen lisääntyy kesää kohden. Näissä klustereissa voi olla kyse Karasekin mallin mukaan (1979) mukaan siitä, että aika ei riitä. Arvioinnin toteuttaminen työn vaatimuksena voisi näin ollen olla yksi kehittämisen kohde.

Edistymisen seuranta ilmiönä voidaan liittää proaktiivisuuteen, jolla tarkoitetaan ennakkointia ja tulevaan varautumista, mikä omalta osaltaan edistää työn vaatimusten hallintaa ja siten työhyvinvointia. Edistymisen seuranta on myös järjestelmätason ominaisuus, joka on osa opettajien Moodle-toimintaa. Jos opettajat voivat valmistautua tulevaan toimintaan, kuten oppilaiden auttamiseen varhaisessa vaiheessa Moodlen avulla, järjestelmää on mahdollisuus suunnitella niin, että entistä parempi proaktiivisuus tulee mahdolliseksi. Edistymisen seurannalla on vuonna 2020 suurempi rooli kuin vuonna 2019. Vuonna 2020 edistymisen seurannan aktiivisuus lisääntyy useissa klustereissa viikon 10 jälkeen, jolloin COVID 19 -pandemian alkoi. Siitä hetkestä lähtien työ kaikilla aloilla, myös opetuksessa siirtyi mahdollisuuksien mukaan virtuaalisille ja digitaalisille alustoille. Kun opetus muuttui etänä tehtäväksi ja kasvokkain tapahtuva vuorovaikutus opiskelijoiden edistymisen tiedon lähteenä katosi, kyky tehdä jotain vastaavaa Moodlessa nousi ajankohtaisemmaksi.

Edistymisen seuranta, arviointi ja Moodlen mahdollisuudet tarvitsevat lisätutkimusta. Edistymisen seuranta voi myös sisältää elementtejä työn vaatimuksista, kun taas Moodlen käyttäminen arvioinnissa voi sisältää elementtejä, jotka parantavat työn hallintaa. Lisätutkimusta tarvitaan myös, jotta voidaan vahvistaa mahdollisuutta tunnistaa asiaankuuluvat interventioiden kohteet. Liittämällä Moodlesta saatuihin tietoihin, kuten klustereihin, terveyteen liittyviä indikaattoreita tai muita tulosmuuttujia, voi olla mahdollista analysoida, miten klusterit eroavat toisistaan esimerkiksi keskimäärin sairauspoissaolojen tai opiskelijoiden lukumäärän suhteen.

5 Päätelmät

Koneoppimismenetelmillä, kuten aihehallinnuksella ja klusteroinnilla, on mahdollista visualisoida työhön liittyvää trace dataa prosesseiksi prosessiteorian ja vaatimukset-hallinta -mallin avulla. Tämä on täysin uusi tapa. Näin ollen voidaan myös vähentää epäilyjä trace datan luotettavasta käytöstä tieteellisesti työhyvinvointitutkimuksen alalla. Järjestelmän lokien digitaalisesta trace datasta on mahdollista löytää ja visualisoida työprosesseja, kun järjestelmää käytetään työn tavoitteiden saavuttamiseen. Aikaleimattu trace data viittaa realistisesti johonkin, joka on suoritettu ajan kuluessa, työn todelliseen sisältöön. Prosessiteorian peruseriaate on, että toiminnan ajallinen eteneminen tulisi olla mahdollista havaita, ja meidän Moodle-tapauksessamme onnistuimme siinä.

Koska Moodlea käytetään 240 maassa (<https://stats.moodle.org>), on kehittämäämme mallia mahdollisuus hyödyntää laajasti Moodle-datan kanssa. Malliamme voi hyödyntää myös muiden työn suorittamiseen käytettävien järjestelmien yhteydessä syntyvään dataan. Selvitystyömme voi avata aivan uusia mahdollisuuksia työhyvinvoinnin kehittämiseen. Havaitsemalla toiminnan ajallista etenemistä voidaan tunnistaa huippuviikot ja aikarajat, joita organisaatiot voivat ottaa kohteiksi omien kehitystarpeidensa näkökulmasta. Tällainen tarve voi olla kuormittavien työn vaatimusten vähentäminen.

Lähteet

- Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003). 'Latent dirichlet allocation', *Journal of Machine Learning Research*, Vol. 3, pp.993-1022.
<https://dl.acm.org/doi/10.5555/944919.944937>
- Elovainio, M. et al. Psychosocial work environment as a dynamic network: A multi-wave Cohort Study. *Scientific Reports* 12, (2022). <https://doi.org/10.1038/s41598-022-17283-z>
- Langley, A. (1999), 'Strategies for theorizing from process data', *Academy of Management Review*, Vol. 24 No. 4, pp.691–710.
<https://doi.org/10.5465/amr.1999.2553248>
- Langley, A., Smallman, C., Tsoukas, H. & Van de Ven, A. H. Process studies of change in organization and management: Unveiling temporality, activity, and flow. *Academy of Management Journal* 56, 1–13 (2013). <https://doi.org/10.5465/amj.2013.4001>
- Lindberg, A. (2020), 'Developing theory through integrating human and machine pattern recognition', *Journal of the Association for Information Systems*, Vol. 21 No. 1, pp. 90-116. <https://doi.org/10.17705/1jais.00593>
- Karasek, R. A. Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly* 24, 285 (1979).
<https://doi.org/10.2307/2392498>
- Landau, S., Leese, M., Stahl, D., & Everitt, B. S. (2011). *Cluster analysis*. John Wiley & Sons. doi.org/10.1002/9780470977811
- Moodle. Moodle statistics, <https://stats.moodle.org> (2022).
- Pentland, B., Vaast, E. and Wolf, J.R. (2021), 'Theorizing process dynamics with directed graphs: a diachronic analysis of digital trace data', *MIS Quarterly*, Vol. 45 No. 2, pp. 967-984. <https://doi.org/10.25300/MISQ/2021/15360>
- Spector, P.E. and Pindek, S. (2016), 'The future of research methods in work and occupational health psychology', *Applied Psychology*, Vol. 65 No. 2, pp. 412-431.
<https://doi.org/10.1111/apps.12056>

Raportissa kerrotaan Työsuojelurahaston rahoittamasta hankkeesta "Organisaatiodatan hyödyntämisen luotettavuuden lisääminen teorian ja tekoälyn avulla". Hanke on teoreettismetodologinen, ja sen tavoitteena oli laatia kaksi kansainvälistä tieteellistä artikkelia. Tutkimusaineisto koostuu kahden ammattikorkeakoulun opettajien Moodle-datasta. Moodle on virtuaalinen oppimisympäristö, jota opettajat käyttävät opetuksen suunnittelua, toteutusta ja arviointia varten. Ensimmäisessä artikkelissa osoitetaan, miten järjestelmien käytössä syntyvä trace data voidaan kiinnittää prosessiteorian viitekehykseen (Langley ym., 2013). Näin ollen tulee mahdolliseksi havaita erilaista toiminnan ajallista etenemistä eli prosesseja ja tehdä päätelmiä työn tekemisen prosessien kuormittavista ajanjaksoista. Päätelmien apuna käytetään Karasekin työn vaatimukset-hallinta -mallia (1979). Toisessa artikkelissa jatketaan analyysiä tarkemmalle tasolle ja käytettiin laskennallisena menetelmänä klusterointia, jolla tunnistetaan kuusi erilaista prosessia tehdä opetustyötä keväällä 2019 ja 2020.



Työsuojelurahasto
Arbetskyddsfonden
The Finnish Work Environment Fund

Työterveyslaitos
Arbetshälsoinstitutet
Finnish Institute of Occupational Health

PL 40, 00032 Työterveyslaitos

www.ttl.fi

ISBN 978-952-391-077-5 (PDF)

